

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2003, 14:12:34 ; Search time 53 Seconds
(without alignments)
6793.180 Million cell updates/sec

Title: US-10-007-521-11

Perfect score: 1174
Sequence: 1 GAGCAGACCCCTCAAGCTG.....AAAAAAAAAAAAAAAAAAAA 1174

Scoring table: IDENTITY_NUC
Gapop 10.0, Capext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued_Patents_NA:*
1: /cgn2_6/pcodata/1/lna/5A_COMB.seq:*
2: /cgn2_6/pcodata/1/lna/5B_COMB.seq:*
3: /cgn2_6/pcodata/1/lna/5A_COMB.seq:*
4: /cgn2_6/pcodata/1/lna/5B_COMB.seq:*
5: /cgn2_6/pcodata/1/lna/PCTUS_COMB.seq:*
6: /cgn2_6/pcodata/1/lna/backfillseq.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1174	100.0	1174	2	US-08-872-437-1 Sequence 1, Appl1
2	1174	100.0	1174	3	US-08-651-136C-11 Sequence 11, Appl1
3	1174	100.0	1174	4	US-09-229-911A-11 Sequence 11, Appl1
4	451.6	38.5	922	4	US-09-189-060B-65 Sequence 65, Appl1
5	437.4	37.3	1060	1	US-08-090-013-1 Sequence 1, Appl1
6	437.4	37.3	1060	1	US-08-081-328-1 Sequence 1, Appl1
7	437.4	37.3	1060	1	US-08-232-249-1 Sequence 1, Appl1
8	437.4	37.3	1060	2	US-08-921-426-7 Sequence 7, Appl1
9	437.4	37.3	1060	2	US-08-833-642A-1 Sequence 1, Appl1
10	437.4	37.3	1060	2	US-08-140-008A-3 Sequence 3, Appl1
11	437.4	37.3	1060	2	US-08-389-423-1 Sequence 1, Appl1
12	437.4	37.3	1060	3	US-08-816-915-7 Sequence 7, Appl1
13	437.4	37.3	1060	4	US-09-189-060B-55 Sequence 55, Appl1
14	437.4	37.3	1060	4	US-09-230-665-1 Sequence 1, Appl1
15	437.4	37.3	1060	4	US-09-189-028-1 Sequence 1, Appl1
16	437.4	37.3	1060	5	PCT-US95-07743-7 Sequence 7, Appl1
17	411.4	35.0	919	4	US-09-189-060B-67 Sequence 67, Appl1
18	392.2	33.4	807	4	US-09-230-665-5 Sequence 5, Appl1
19	392	33.4	922	4	US-09-189-060B-73 Sequence 73, Appl1
20	377.2	32.1	1257	3	US-09-230-222-2 Sequence 2, Appl1
21	375.6	32.0	1257	4	US-09-230-225B-5 Sequence 5, Appl1
22	369.8	31.5	1154	3	US-08-651-136C-7 Sequence 7, Appl1
23	369.8	31.5	1154	4	US-09-229-911A-7 Sequence 69, Appl1
24	364.6	31.1	924	4	US-09-189-060B-69 Sequence 4, Appl1
25	364	30.1	927	4	US-09-254-733-4 Sequence 30, Appl1
26	353.6	28.7	894	3	US-08-651-136C-3 Sequence 3, Appl1

28	337	28.7	894	4	US-09-229-911A-3 Sequence 3, Appl1
29	336.2	28.6	912	4	US-09-189-060B-71 Sequence 71, Appl1
30	334.8	28.5	927	3	US-08-651-136C-5 Sequence 5, Appl1
31	334.8	28.5	927	4	US-09-229-911A-5 Sequence 5, Appl1
32	315.2	26.8	1473	1	US-08-090-013-3 Sequence 3, Appl1
33	315.2	26.8	1473	1	US-08-081-328-3 Sequence 3, Appl1
34	315.2	26.8	1473	1	US-08-232-249-3 Sequence 3, Appl1
35	315.2	26.8	1473	2	US-08-833-642A-3 Sequence 3, Appl1
36	315.2	26.8	1473	2	US-08-389-423-3 Sequence 3, Appl1
37	315.2	26.8	1473	4	US-09-230-665-3 Sequence 3, Appl1
38	315.2	26.8	1473	4	US-09-189-028-3 Sequence 3, Appl1
39	313.6	26.7	885	3	US-08-651-136C-23 Sequence 23, Appl1
40	313.6	26.7	885	4	US-09-229-911A-23 Sequence 23, Appl1
41	302	25.7	1132	3	US-08-651-136C-21 Sequence 21, Appl1
42	302	25.7	1132	4	US-09-229-911A-21 Sequence 21, Appl1
43	288.2	24.5	1423	3	US-08-651-136C-9 Sequence 9, Appl1
44	288.2	24.5	1423	4	US-09-229-911A-9 Sequence 9, Appl1
45	286.2	24.4	964	4	US-09-230-665-7 Sequence 7, Appl1

ALIGNMENTS

RESULT 1
US-08-872-437-1
; Sequence 1, Application US/08872437
; Patent No. 5958082
; GENERAL INFORMATION:
; APPLICANT: Lund, Henrik
; APPLICANT: Kalum, Lisbeth
; TITLE OF INVENTION: Garments With Considerable Variation In
; TITLE OF INVENTION: Abrasion Level
; FILE REFERENCE: 4888-200-US
; CURRENT APPLICATION NUMBER: US/08/872.437
; EARLIER FILING DATE: 1997-06-10
; EARLIER APPLICATION NUMBER: 1276/96
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1174
; TYPE: DNA
; ORGANISM: Thielavia terrestris
US-08-872-437-1

Query Match 100.0%; Score 1174; DB 2; Length 1174;
Best Local Similarity 100.0%; Pred. No. 5.7e-224;
Matches 1174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	GAGCAGACCCCTCAAGCTGTTCCACCCCGCTCTTTTCGGCCCGCAGGA	60
DB	1	GAGCAGACCCCTCAAGCTGTTCCACCCCGCTCTTTTCGGCCCGCAGGA	60
QY	61	TGCGCTTACTCCGCTTTCGCAACCCGCGCTGCACTCTCTGTCGCTCG	120
DB	61	TGCGCTTACTCCGCTTTCGCAACCCGCGCTGCACTCTCTGTCGCTCG	120
QY	121	CGGCAATGCGAGTGGCAGTCCAGAGATCTGGAGCTGTCGAAGCCGTGCGCTT	180
DB	121	CGGCAATGCGAGTGGCAGTCCAGAGATCTGGAGCTGTCGAAGCCGTGCGCTT	180
QY	181	GGCCGGGGAAGCGCGGTGACCAACCGGCTGCGGTGATGCCAATTCAGAGGCC	240
DB	181	GGCCGGGGAAGCGCGGTGACCAACCGGCTGCGGTGATGCCAATTCAGAGGCC	240
QY	241	TGTCGACTTCAATGTCAAGTGGGCTGCAACGGCGGCTGCTACTCTGCGCAGAC	300
DB	241	TGTCGACTTCAATGTCAAGTGGGCTGCAACGGCGGCTGCTACTCTGCGCAGAC	300
QY	301	AGATCCCTGCGCGGTGAAGACAAATCTGCGCTTCCGCGGACAGCATTCGCGG	360
DB	301	AGATCCCTGCGCGGTGAAGACAAATCTGCGCTTCCGCGGACAGCATTCGCGG	360

[illegible]

	?	ADDRESS:	No. 60016390 No. 6001639disk of No. 6001639th America, Inc
	?	STREET:	405 Lexington Avenue, 64th Floor
	?	CITY:	New York
	?	STATE:	New York
	?	COUNTRY:	United States of America
	?	ZIP:	10174-6401
	?	COMPUTER READABLE FORM:	
	?	MEDIUM TYPE:	Floppy disk
	?	COMPUTER:	IBM PC compatible
	?	OPERATING SYSTEM:	PC-DOS/MS-DOS
	?	SOFTWARE:	Patentlin Release #1.0, Version #1.30
	?	CURRENT APPLICATION DATA:	
	?	APPLICATION NUMBER:	US/08/651,136C
	?	FILING DATE:	21-MAY-1996
	?	CLASSIFICATION:	435
	?	ATTORNEY/AGENT INFORMATION:	
	?	NAME:	Lambiris, Elias J.
	?	REGISTRATION NUMBER:	33,728
	?	REFERENCE/DOCKET NUMBER:	4366-.200-US
	?	TELECOMMUNICATION INFORMATION:	
	?	TELEPHONE:	212-867-0123
	?	TELEFAX:	212-878-9655
	?	INFORMATION FOR SEQ ID NO:	11:
	?	SEQUENCE CHARACTERISTICS:	
	?	LENGTH:	1174 base pairs
	?	TYPE:	nucleic acid
	?	STRANDEDNESS:	single
	?	TOPOLOGY:	linear
	?	MOLECULE TYPE:	CDNA
	?	FEATURE:	
	?	NAME/KEY:	CDS
	?	LOCATION:	60..956
	?	US-08-651-136C-11	
		Query Match	100.0%; Score 1174; DB 3; Length 1174;
		Best Local Similarity	100.0%; Pred. No. 5.7e-224;
		Matches 1174; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
OY	1	GAGCAGCACCCCTCAAGCTGACGTACATTTCCACC	CGGCTCTCTTTTCTGGCCCCCAGA 60
Db	1	GAGCAGCACCCCTCAAGCTGACGTACATTTCCACC	CGGCTCTCTTTTCTGGCCCCCAGA 60
OY	61	TGCAGCTACTGCCGTCCTTCGCACAACCC	TGGCGCTGCATTCCTCTGTCGCTCG 120
Db	61	TGCAGCTACTGCCGTCCTTCGCACAACCC	TGGCGCTGCATTCCTCTGTCGCTCG 120
OY	121	CGGCCAGTGCGCAGTGGCCAGTCCACGAGATCTGGGACTGTGCAAGCCGTCGTGGCTT	180
Db	121	CGGCCAGTGCGCAGTGGCCAGTCCACGAGATCTGGGACTGTGCAAGCCGTCGTGGCTT	180
OY	181	GGCCCCGGGAAGGCGCCGTCGACCAACCGGTTACGCGTGCATGCGCACTTCCAGCGCC	240
Db	181	GGCCCCGGGAAGGCGCCGTCGACCAACCGGTTACGCGTGCATGCGCACTTCCAGCGCC	240
OY	241	TGTCCGACTTCAAATGCCAGTGGGGCTGGCAAGGGGGGCTCGGCGCTTACTCTCGCGGACC	300
Db	241	TGTCCGACTTCAAATGCCAGTGGGGCTGGCAAGGGGGGCTCGGCGCTTACTCTCGCGGACC	300
OY	301	AGACTTCCTGGCGGGTGAAGCACAATCTCGCCTACGCGCTTCCGCGGAGAGCATCGCG	360
Db	301	AGACTTCCTGGCGGGTGAAGCACAATCTCGCCTACGCGCTTCCGCGGAGAGCATCGCG	360
OY	361	CGGGTCCGGAATCTCGTGGTGTCTGCGGCTGCTACGCGCTCACTTCACTTCCGCTCCG	420
Db	361	CGGGTCCGGAATCTCGTGGTGTCTGCGGCTGCTACGCGCTCACTTCACTTCCGCTCCG	420
OY	421	TGCGCGGCAAGCAATGTGTGTGCAAGTCAACGACATGTGGCGGCACTGGGAAGTAAC	480
Db	421	TGCGCGGCAAGCAATGTGTGTGCAAGTCAACGACATGTGGCGGCACTGGGAAGTAAC	480
OY	481	AGTTGATATCGCATTCGCCCGGCGGCGGCGTGGGATCTTCAAGGCTCAGCTCGCAGT	540
Db	481	AGTTGATATTCGCCATTCGCCCGGCGGCGGCGTGGGATCTTCAAGGCTCAGCTCGCAGT	540

COUNTRY: USA
ZIP: 45217
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/090,013
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA: EP 91202880.0
APPLICATION NUMBER: EP 91202880.0
FILING DATE: 06-NOV-1991
ATTORNEY/AGENT INFORMATION:
NAME: PATEL, KEN K.
REGISTRATION NUMBER: 33,988
REFERENCE/DOCKET NUMBER: CM393
TELECOMMUNICATION INFORMATION:
TELEPHONE: 513-627-6437
TELEFAX: 513-627-4854
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1060 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
FEATURE:
NAME/KEY: CDS
LOCATION: 10..924
US-08-090-013-1

Query Match 37.3%; Score 437.4; DB 1; Length 1060;
Best Local Similarity 70.4%; Pred. No. 3e-78;

Matches 655; Conservative 0; Mismatches 246; Indels 30; Gaps 4;

QY 55 CCAGATGCGCTTACTCCGCTTCTTGACCAACCTGGCCGCTGCACTTCTGTGTCG 114
DB 5 CCAAGATGCGCTTCTCTCCCTCCCTCCGCTTGTGGCCGCTGCTGTG 64
QY 115 CCGCGGCGGCGCATGCGAGTGGCAGTCCAGAGATCTGGAGCTGTGCAAGCCGTCGT 174
DB 65 CCGCTTGGCGTGTAGGAG-----GTCCACCGCGTACTGGAGTGTGCAAGCCGTCGT 118
QY 175 GCGCTTGGCCGCGGAGGCGCGCTGACAGCAACCGGCTGACGCTGCGATGCACTTCC 234
DB 119 GCGGCTGGCGCAAGAGGCTCCGCTGACCAAGCTGTCTTCTGCAAGCCGCACTTCC 178
QY 235 AGCGCTGTCCGACTTCAATGTCCAGTGGGCTGCAA--CGCGGCTCGGCTTACTCT 291
DB 179 AGCGTATCAGGAGTTCGACGCAAGTCCGCGTGGAGCGGCGGTGTGCGCTACTCGT 238
QY 292 GCGCGGCGGCGAGCTCCCTGGGCGGTGAAGCAATCTGCGCTAGCGCTTGGCCGCA 351
DB 239 GCGCGGCGGCGAGCGCGGCTGTGAGCAGCACTTGGCTGCGTGGTGGTGGCACT 298
QY 352 GCATGCGCGGCGGCGGCTGCAATCTGCTGCTGCGCTGCTGCAAGCGCTGCACTTCACTT 411
DB 299 CTAATGCGCGCAGCAATGAGGCGGCGTGTGCTGCGCTGCTGCAAGCGCTGCACTTCACT 358
QY 412 CCGGCTCGCGCGCGGCGGCAAGCAATGCTGCTGCAAGCAGCAGTGGCGGCGCACTGG 471
DB 359 CCGGCTCGCTGTGCGGCAAGCAATGCTGCTGCAAGCAGCAGTGGCGGCGTATCTTG 418
QY 472 GAAGTAAACGATGCGATATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 531
DB 419 GCAGCAACCACTTGTGATCTCAACATCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 478
QY 532 GCTCGGAGTTCGCGGCGGCTCCCGGCTCAATAGCGGCGGCGGCGGCGGCGGCGGCGG 591
DB 479 CTCGCCAGTTCGCGGCGGCTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 538
QY 592 GCGATTCCTTCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 651

DB 539 GCGATGCGTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 598
QY 652 AGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 711
DB 599 AGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 658
QY 712 GCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 771
DB 659 GCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 718
QY 772 GCAACG-----TGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 813
DB 719 GCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 778
QY 814 GCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 870
DB 779 CGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 838
QY 871 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 930
DB 839 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 898
QY 931 TGAAGCACTTACTGCGCAAGTGCCTTAAC 961
DB 899 TTAATGACTGTGATCATCATGCTGTAGAC 929

RESULT 6

US-08-081-328-1
Sequence 1, Application US/08081328
Patent No. 5520838

GENERAL INFORMATION:
APPLICANT: BAECK, ANDRE C.

APPLICANT: CEULEMANS, RAPHAEL ANGELENE A.
TITLE OF INVENTION: COMPACT DETERGENT COMPOSITIONS WITH
TITLE OF INVENTION: HIGH ACTIVITY CELLULOSE
NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:
ADDRESSEE: THE PROCTER & GAMBLE COMPANY
STREET: 11810 East Miami River Road
CITY: CINCINNATI
STATE: OHIO
COUNTRY: USA
ZIP: 45253-8707

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/081,328
FILING DATE: 11/19/93
CLASSIFICATION: 252
ATTORNEY/AGENT INFORMATION:
NAME: ZERBY, KIM WILLIAM
REGISTRATION NUMBER: 32,323
REFERENCE/DOCKET NUMBER: CM356M
TELECOMMUNICATION INFORMATION:
TELEPHONE: 513-627-2885
TELEFAX: 513-627-0318
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1060 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
FEATURE:
NAME/KEY: CDS
LOCATION: 10..924
US-08-081-328-1

Query Match	Similarity	70.4%	Score 437.4	DB 1	Length 1060
Best Local Similarity	70.4%	Pred. No. 3e-78			
Matches 655	Conservative	0	Mismatches 246	Indels 30	Gaps 4
QY	55	CCAGAGATGCGCTTACTCCGCTTCTTCGCACAAACCCCTGGCCGCTGCACCTTCTGGTGG	114		
Db	5	CCAAGATCGTTCCTCCGCCCTCCGCCGCTGGCTGGCCGCGCCCTGCCGGGTGGT	64		
QY	115	CCTCCGCGCCAGTGGCAGTGGCCAGTCCACGATTACTGGGACTGCTGCAAGCCGTGCT	174		
Db	65	CCCTTGGCCGCTGATGGCAG-----GTCCACCCTGCTACTGGGAGTGGCAAGCTTCTCT	118		
QY	175	GCGCTTGGCCCGGGGAAGGCGCGCTGCAGGCAACGCGGTTCAGCGGCGGATGGCAACTTCC	234		
Db	119	GCGGCTGGGCCCAAGAGGCTCCCGTGAACCAACGCTGTCTTTCCTGCAACGCAACTTCC	178		
QY	235	AAGCGCTTCCGCACTTCATGTCCAGTGGGGCTGCAA--CGCGGGCTTCGGCTACTCTCT	291		
Db	179	AAGCATATCAGGACTTCGACGCGCAAGTCCGGCTGGAGCGCGGGCGGCTGTCCGCTACTGCT	238		
QY	292	GCGCGGACCACTCCCTCGGGGCGGTGAACGACAAATCTCGCTTACGCGCTTCGCGCGAGCA	351		
Db	239	GCGCGGACCAAGCCCATATGGGCTGTAACAGACACTTGCCTCGCTTTCGCTGCACCT	298		
QY	352	GCATGCGCGCGGGGCTCCGATTCCTGTGTGTGTGGCGCTGCTACGCGCTCACCCTTCACTT	411		
Db	299	CTATTGTCGGGCAAGCAATAGGGGGGCTGGTGGCGCTGCTACGACGCTCACCTTTCACAT	358		
QY	412	CCGGTCCCGCGCGCCGCAAGACATGTTGGTGCAGTCAACGACAGCACTGGCGCGCACTCG	471		
Db	359	CCGGTCTCTGTGTGTGGCAAGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	418		
QY	472	GAAGTAAACCACTTGCATATCGCATCCCGCGCGGGCGCTGGGCAATCTTCAACGCGTGA	531		
Db	419	GCAGCAACCACTTGCATCTCAATCCCGCGCGGGCGCTCGGCACTTCGACGAGATGA	478		
QY	532	GCTGCCACTTGGGGGGGCTCCCGCGCGCTCAATACGGCGGCAATTCGTGGCCGAGCACT	591		
Db	479	CTCCCACTTGGGGGGGCTGTGGCGCGCGCAAGGCTACGGCGGCAATCTGTGCGCAAGAT	538		
QY	592	GCGATTTCCTCCCGCGCGGCTCAAGACCCGCGCTGCACAGTGGAGCGGTTTGACTGGTTCAGA	651		
Db	539	GCGATTCGTTCCCGCACGCGCTCAAGCCCGGCTGCTACTAGGGGCTTCGACTGGTTCAGAA	598		
QY	652	ACGCGGACAAACCGAGTTTACGTTTCCAGACAGTGGCAGTGGCCCGCGGAGATGTGGCC	711		
Db	599	ACGCGGACAAATCCGAGCTTTCAGCTTCCGTCAGAGTCCAGTGGCCAGCGCGAGCTGTGCTGC	658		
QY	712	GCTCGGCTGCAAGCGGCAACGACGACTCCAGTTCCTCCCGCTTCACACCCCGCAAGGGGTG	771		
Db	659	GCACGGGATGTCGCGCGCAACGACGCGCAATCTTCCTTCGCGTCCAGATCCCTCCAGAA	718		
QY	772	GCACGGG-----TGCGACCGGAGCGCCACAGTCCAGTGGCGCTGGGT	813		
Db	719	GCACCAAGCTTTCGCGTCAACGAGCTACCAAGACCAAGCAACAGTCCACTCCACACACT	778		
QY	814	CGGCGCAAGC--GTCTCCCGGGCGGCGGCTGCAGTGGCTGCAGTCTCAAGATGGGCTCACT	870		
Db	779	CGAGCCCGCAATCCAGCTTACGACTCCACAGCGGCTGCACTGCTGAGAGTGGGCTCACT	838		
QY	871	GCGGTGGCATCGGCTTTCAGCGGATGACACACTGTGTCTCTGGACACACTCGCAGAAAT	930		
Db	839	GCGGGGCAATGTGGTGGAGCGGCTGCACACCTGGGTGCTGGGCACTTTCGACAGAA	898		
QY	931	TGAACGACTACTGCTGCAGTGCCTCTTAAC	961		
Db	899	TTAATGACTGTTACCATCATGATGCTGTAGAC	929		

```

GENERAL INFORMATION:
APPLICANT: MCCORDDALE, FINLAY (NMN)
APPLICANT: BUSCH, ALFRED (NMN)
TITLE OF INVENTION: DYE TRANSFER INHIBITING COMPOSITIONS
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: THE PROCTER & GAMBLE COMPANY
STREET: 5299 SPRING GROVE AVENUE
CITY: CINCINNATI
STATE: OHIO
COUNTRY: USA
ZIP: 45217
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,249
FILING DATE: 05-MAY-1994
CLASSIFICATION: 510
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 91202892.6
FILING DATE: 06-NOV-91
ATTORNEY/AGENT INFORMATION:
NAME: ALLEN, GEORGE W.
REGISTRATION NUMBER: 26,143
REFERENCE/DOCKET NUMBER: CM-395
TELECOMMUNICATION INFORMATION:
TELEPHONE: 513-627-5946
TELEFAX: 513-627-8118
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1060 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
FEATURE:
NAME/KEY: CDS
LOCATION: 10..924
US-08-232-249-1

Query Match      37.3%; Score 437.4; DB 1; Length 1060;
Best Local Similarity 70.4%; Pred. No. 3e-78;
Matches 655; Conservative 0; Mismatches 246; Indels 30; Gaps

QY    55 CCAGGATGGCGTCTACTCCCGTTCTTCGCACACCCTGGCCGCTGCACCTTCTGTGGTG 114
Db     5 CCAAGATGGGTTCCTCCCCCTCCGCCGCTTGTGGCCGCGCTGCGGTTGG        64
QY    115 CCTCCGGCGCAGTGGCAGTGGCAGTGCACAGATACAGAGATCTGGGAGTCTCAGACCGCTGCT 174
Db     65 CCCTTGGCCGCTGATGGCAG-----GTCCACCCGCTACTGAGGAGCTGTCGAAGCCTTCTGT 118
QY    175 GGCGTTGGGCGCGGAAGGCGCGCGTCAAGCACACCGGTGTACCGTGNCGCATGTGCACACTTGC 234
Db    119 GGCGGTGGGCCAAGAAGGCTCCCCTGTAACACAGCTGTCTTTCTTCGTGCAAGCGCACTTGC   178
QY    235 AGCGCGTGTCCGACCTTCATATGTCAGATCGGCGCTGCAA--CGGCGGCTTGGCGCTACTCTCT 291
Db    179 ACGCATATCAGGACCTTCGACGCGCAGATGCGCTGGGAGCGGCGGCTGTGCGCTACTCTGT   238
QY    292 GCGCCGACGACACTCCCTGTGGGCGGTGAACGACACAATCTCGCCTAACGGCTTCCGCGGAGACA 351
Db    239 GCGCGCAGACAGACCCCATATGGGCTGTGAACGACGACTTCCGCGCTTTTGTGTCACCT       298
QY    352 GCATGGCGCGGCGGATCCGAAATCCTGTGTGGTGTGGGCGCTTACGCGCTCACCTTCACTT    411
Db    299 CTATTGGCGCGCAGCATATGGGCGGCGCTGTGTGCGCTGTCAAGACTTCACTTACAT    358
QY    412 CGGGTCCGCGTCCCGCGCAGACATATGTTGTGTACAGTCAACGACGACTGGCGGAGCACTGG 471
Db    359 CGGGTCTGTCTGTGGCAGAAGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 418

```

QY	472	GAATTAACAGCTTGGATATGCGATCCATCCCGGCGGGCGCTGGGACATCTTCAACGGCTGCA	531
Db	419	GCAGCAACCACTTGATCTCAATCAATCCCGCGGGCGCTGGGACATCTTCAACGGATGCA	478
QY	532	GCTTCGCACTTCGGCGGCGCTCCCGGCGCTCAATACGGCGGCAATTTCGTGCGCGACAGT	591
Db	479	CTCCCCAGTTTCGGGGGGTCTCCCGGCGAGGGCTACGGCGGCAATCTCTCCCGCAACGAGT	538
QY	592	GGATATTCCTTCCCGCGGCGCTCAAGCCCGGCGTGGCAGTGGGGGTTTGACTGCTTCCAGA	651
Db	539	GGGATTCGGTTTCCCGGAGCGCTTCAAGCCCGGCTGCTACTTGGCGCTTCGACTGTGTTCAAGA	598
QY	652	ACGCGCAACACCGGACGTTTACGTTCCAGAGGTGCAGTGGCCCCGCGGAGTCTGTGCCC	711
Db	599	ACGGCGCAATTCGGAGCTTCAAGTCTCCGTTCCGTTGAGTTCAGTGGCCAGCGGACGAGCTGTGCTC	658
QY	712	GCTCCGGCTGCAAGCGCAAGGAGGACATCCAGTCTCCGCTTTCACACCCCGCAAGCGGGT	771
Db	659	GCACCGGATGCGCGCGGCAACGACGAGCGGCAACTTCCCTGCGTTCAGATCCCTTCAGCA	718
QY	772	GCAACGG-----TGGCACCGGGAGCGCCACGTCGTGCGCTGGGT	813
Db	719	GCACACACTCTCCGCTCAACCACTTACAGGACACGAGCAACAGTCACTTCACACACT	778
QY	814	CGGGGCAAGAC---GTCTCCGGGGGGGGGAGAGTGGCTGCAACGTCTGAGAGTGGGCTCACT	870
Db	779	CGAAGCCCGGCAATCCAGCTTACGACTTCCACGCGGCTGCACTCTGTGAGAGTGGGCTCACT	838
QY	871	GGCGTGGCACTGGCTTACGGGATGACACCACTGTCTTCTTGGACACACACTCCAGAGAT	930
Db	839	GGCGGGCAATGGCTGTGAGGCGCTGCACCACTGCGTGGCGGACACTTGCACGAGA	898
QY	931	TGACGACACTACTCTGCACTGCCTCTTAAC	961
Db	899	TTAATGACTGGTACCATCACTAGTCTGTGAGAC	929

RESULT 8
 US-08-921-426-7
 Sequence 7, Application US/08921426
 Patent No. 5837847
 GENERAL INFORMATION:
 APPLICANT: Royer, John C
 APPLICANT: Moyer, Donna L
 APPLICANT: Yoder, Wendy T
 APPLICANT: Shuster, Jeffrey R
 TITLE OF INVENTION: NON-TOXIC, NON-TOXIGENIC, NON-PATHOGENIC
 TITLE OF INVENTION: FUSARIUM EXPRESSION SYSTEM
 NUMBER OF SEQUENCES: 16
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: NO. 5837847/0 No. 5837847/disk of No. 5837847/1h America, Inc.
 STREET: 405 Lexington Avenue, 64th Floor
 CITY: New York
 STATE: New York
 COUNTRY: USA
 ZIP: 10174-6401
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/921,426
 FILING DATE: 29-AUG-1997
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/456,433
 FILING DATE: 01-JUN-1995
 APPLICATION NUMBER: US 08/404,678
 FILING DATE: 15-MAR-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Agtis Dr., Cheryl H.

```

?      REGISTRATION NUMBER: 34,086
?      REFERENCE/DOCKET NUMBER: 4216.010-US
?      TELECOMMUNICATION INFORMATION:
?      TELEPHONE: 212-867-0123
?      TELEFAX: 212-878-9655
?      INFORMATION FOR SEQ ID NO: 7:
?      SEQUENCE CHARACTERISTICS:
?      LENGTH: 1060 base pairs
?      TYPE: nucleic acid
?      STRANDEDNESS: single
?      TOPOLOGY: linear
?      FEATURE:
?      NAME/KEY: CDS
?      LOCATION: 10..924
?      FEATURE:
?      NAME/KEY: mat_peptide
?      LOCATION: 73..924
?      FEATURE:
?      NAME/KEY: sig_peptide
?      LOCATION: 10..72
?      US-08-921-426-7

37.3%; Score 437.4; DB 2; Length 1060;
Best Local Similarity 70.4%; Pred. No. 3e-78;
Matches 655; Conservative 0; Mismatches 246; Indels 30; Gaps 4

```

QY	55	CCAGAGTACGGCTTCAATCCGCTTTTGCGCAACACCCGCGCGGACACTTCCTGGTGC	114
Db	5	CCAGATATCCGTTCTCCCTCCCTCCTCCCGTGGCGCTTGTGGCGCGCTCCGGTGTGG	64
QY	115	CTCTCCGGGGCGATGGCGCATGGGCCATGCCAGAGATATCTGGAGCTGCTGCAGACCGTGT	174
Db	65	CCCTTGGCGCGCTGATGGAGAG-----GTCCACCCCGCTACTGGAGATGCTGCAGACCTTTCGT	118
QY	175	GGCGTTGGCCCCGGGAAGAGCCGCGTCAGACCAACCGGACTAGCGGCTGGAGTGGCCAACTTTC	234
Db	119	GGCGCTGGGGCCAGAGAGGCTCCGCTGAACACGCTGTCTTTTCTGTGCAAGGCCAACTTTC	178
QY	235	AGCGCTGTGTCCGACTTCAATGTCCAGTGGGGCTGCAG--CGGGGCTCGGCGCTTACTCT	291
Db	179	AGCGTATCAACGAGACTTGCAGCGCCAAAGTCCGGCTCGAGACGGGGGGGTGGCGCTACTCGT	238
QY	292	GGCGCGACACACATCCCGTGGGGGCTGAAACACAAATCAGCTTACGGCTTGGCCCGGACGA	351
Db	239	GGCGCGACACACACCCCATGGGCTGTGAACGACGACTTTCGGCTGGCTTGTGTCCACCT	298
QY	352	GCATCGCGCGGGGTCCGAAATCTCTGTGAGTCTCGGCTCGACTACGCGCTACCTTCACTT	411
Db	299	CTATTGGCGGGACGAGAAATGAGCGGGGTGTGTGTGCTGGCTGTAGAGACTCACTTCAAT	358
QY	412	CCGGTCCCGCTGCGCGCACAGCAATGTGTGTGCAATCAGACAGCACTGGCGGCGACCTGG	471
Db	359	CCGGTCCGTGTGCTGGCAGAGAAATGTGTCTGCAGTCCACAGCACTGGGGGTGATCTTGG	418
QY	472	GAAAGAACCACTTGTGATATCCATTCGCCCGCGCGCGCGCGGCGGCGATTTCAAGGGCTGCA	531
Db	419	GCAGCAACCACTTGTGATCTCAACATCCCCCGCGCGCGCGCTGCGCATTTTGACAGGATGCA	478
QY	532	GCTCGCAATTTGGGGGCGCTCCCGCGGCGCTCAATACGGCGGCAATTTCTGCGCGGACCAAT	591
Db	479	CTTCCCAATTTGGGGGCGTCTGCCCGGACAGCGCTACGGCGGCAATCTCTGCCGCAACGAGT	538
QY	592	GGCATTCCTTCCCGCGCGCGCTGCAAGCCCGGCTCCAGTGGCGTTTGATCTGTTTCAGA	651
Db	539	GGCATCGGTTCCCGGAGCGCTTCAAGCCCGGCTCTACTGGCGCTTGTGACTGTGTTCAAGA	598
QY	652	ACGGCGAACACCCGACGTTCTACGTTTCCAGCAGGTGAGTGCCTCCCGCGAGATCGTTGGCCC	711
Db	599	ACGGCGAACAAATCCGAGGTTACGTTCTTCGTACAGTCCAGTGGCCAGCGGACCTCGTGCCTC	658
QY	712	GCTCGGCTGCAAGCGCACAGACGACTTCCTCCCGTCTTACCCCGCCCAAGGGGTG	771
Db	659	GCACCGGATTTGGCGCGCGCAAGACGAGCGGCAAACTTTCCTTGGCGTCCAGATCTCTTCCAGCA	718

OY 772 GCAACG-----TGACCGGGAGCCGACGTCGACGCGCTGGGT 813
 Db 719 GCACCACTCTCCGGTACACCAAGCTACACGACACCAAGCTCCACACT 778
 OY 814 CGGCGCAGAC--GTCTCCGGGGGGGCGAGTGGCTGCAGCTTCAGAAAGTGGCTCACT 870
 Db 779 CGAGCCCGCAGTCCAGCTACGACATCCGAGCGGCTGCATGCTGAGAGTGGCTCACT 838
 OY 871 GCGGTGCAATCGGTTGACGGGATGACACCACTGTCTGTCGGACCACTGCCAGAGT 930
 Db 839 GCGCGGCAATGCGTGGAGCGGCTGCACCACTGCTGCTGGCAGCACTTCACAGAGA 898
 OY 931 TGAACGACTACTCCGACGTGCTTAAC 961
 Db 899 TTATGACTGTACATCAGTGCCTGTAGAC 929

RESULT 9

US-08-833-642A-1

; Sequence 1, Application US/08833642A

; Patent No. 5883066

; GENERAL INFORMATION:

; APPLICANT: Ivan M. A. J. Herbots et al.

; TITLE OF INVENTION: Liquid Detergent Compositions

; NUMBER OF SEQUENCES: 5

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Jackie Ann Zurcher

; ADDRESS: Dinsmore & Shohl LLP

; STREET: 255 E. Fifth Street

; CITY: Cincinnati

; STATE: Ohio

; COUNTRY: USA

; ZIP: 45202

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.5 inch

; OPERATING SYSTEM: MS-DOS

; SOFTWARE: Wordperfect 6.1

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/833,642A

; FILING DATE: April 8, 1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Zurcher, J. A.

; REGISTRATION NUMBER: P42,251

; TELEPHONE: (513) 977-8377

; TELEFAX: (513) 977-8141

; INFORMATION FOR SEQ ID NO: 1:

; LENGTH: 1060 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

; HYPOTHETICAL: NO

; ORIGINAL SOURCE:

; ORGANISM: Humicola insolens

; STRAIN: DSM 1800

; FEATURE:

; NAME/KEY: mat peptide

; LOCATION: 73..927

; NAME/KEY: s19 peptide

; LOCATION: 10..12

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 10..927

; US-08-833-642A-1

Query Match 37.3%; Score 437.4; DB 2; Length 1060;
 Best local similarity 70.4%; Pred. No. 3e-78;
 Matches 655; Conservative 0; Mismatches 246; Indels 30; Gaps 4;

OY 55 CCAAGATGCGCTCTACTCCGCTTCTTGCAACACCGGGCGGCTGCACTTCTCGGTG 114
 Db 5 CCAAGATGCGCTTCTCCCGCTTCTTGCAACACCGGGCGGCTGCACTTCTCGGTG 64
 OY 115 CTTCCGGGCGCACTGGCAGTGGGCGGAGTCCAGAGATCTGAGGATGTCGAAGCGCTG 174
 Db 65 CCGTTCGCGCTGATGAGAG-----GTCCACCGCGCTACTGGGATGTCGAAGCGCTG 118
 OY 175 GCGCTTGGCCGGGAGAGCGCGGTCCAGCAACCGGTCTACGCTGCGAGTCCAACTTC 234
 Db 119 GCGGCTGGGCGCAAGAGGCTTCCGTGAACAGCGCTGCTTCTTCTGCAACCGCACTTCC 178
 OY 235 AGCGCGCTGCGCACTTCAATGTCAGCGGCTGCA-----CGGGGGCGGCGCTACTGCT 291
 Db 179 AGCGTATCACGAGCTTTCAGCGCAAGTCCGCGTGGAGCGGGGCTGCTACTGCT 238
 OY 292 GCGCGGCGAGACTCCCTGGGGGTGAGCAACATCTCGCTTACGCTTCCGCGAGCA 351
 Db 239 GCGCGGCGAGACTCCCTGGGGGTGAGCAACATCTCGCTTACGCTTCCGCGAGCA 298
 OY 352 GCATCGCGGCGGCGGCTCCGAATCTCTGTGTGCTGCGCTGCTACGCGCTACCTTCACTT 411
 Db 299 CTATTCGCGGCGAGCAATGAGCGGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 358
 OY 412 CCGGTCGCGGCGGCGGAGCAACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 471
 Db 359 CCGGTCGCTGTGCTGCGGAGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 418
 OY 472 GAAGTAACAGTTCGATATGCGCATGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 531
 Db 419 GCAAGCAACACTTTCGATCTCAATCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 478
 OY 532 GCTGCGAGTTGCGGCGGCTCCCGGCGCTCAATACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 591
 Db 479 CTCCCGAGTTGCGGCGGCGGCTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 538
 OY 592 GCGATTCCTTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 651
 Db 539 GCGATTCGTTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 598
 OY 652 ACGCCGCAACACCGAGCTTCAGTTCAGGATGCAAGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 711
 Db 599 AGCGCGCAATTCGAGCTTCAGTTCAGGATGCAAGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 658
 OY 712 GCTCCGCTGCAAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 771
 Db 659 GCAACGGATCCCGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 718
 OY 772 GCAACG-----TGACCGGGAGCCGACGTCGACGCGCTGGGT 813
 Db 719 GCACCACTCTCCGGTACACCAAGCTACACGACACCAAGCTCCACACT 778
 OY 814 CGGCGCAGAC--GTCTCCGGGGGGGCGAGTGGCTGCAGCTTCAGAAAGTGGCTCACT 870
 Db 779 CGAGCCCGCAGTCCAGCTACGACATCCGAGCGGCTGCATGCTGAGAGTGGCTCACT 838
 OY 871 GCGGTGCAATCGGTTGACGGGATGACACCACTGTCTGTCGGACCACTGCCAGAGT 930
 Db 839 GCGCGGCAATGCGTGGAGCGGCTGCACCACTGCTGCTGGCAGCACTTCACAGAGA 898
 OY 931 TGAACGACTACTCCGACGTGCTTAAC 961
 Db 899 TTATGACTGTACATCAGTGCCTGTAGAC 929

RESULT 10

US-08-140-008A-3

; Sequence 3, Application US/08140008A

; Patent No. 5914306


```
FEATURE:
; NAME/KEY: sig-peptide
; LOCATION: 10..72
US-08-816-915-7

Query Match      37.3%; Score 437.4; DB 3; Length 1060;
Best Local Similarity 70.4%; Pred. No. 3e-78;
Matches 655; Conservative 0; Mismatches 246; Indels 30; Gaps 4;

55 CCGAGATGCGCTTACTCCGCTTCTTCGCAACACCCGCGCGCTGCACTTCCTGTGCG 114
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5 CCAAGATGCGCTTCTCCCTCCCTCCCTCCGCTGCGCGCTGCGCGCTGCGTGG 64

OY 115 CCTCCGCGCGCAGTGGAGTGGCCAGTCCAGAGATCTGGAGTGTGCAAGCGTCT 174
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 65 CCTTGGCGCGTGTATGGAG-----GTCCACCCGCTACTGTGGAGTGTGCAAGCGTCT 118

OY 175 GCGCTTGGCGCGGAGAGCGCGCGTCAAGCAACCGGTCTAGCGGTGGATGCCAATTCC 234
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 119 GCGGCTGGCGCAAGAGAGCTCCGCTGAACCAAGCTGTCTTCTTCGCAACGCCAATTCC 178

OY 235 AGGCGCTTCCGACTTCATATGTCAGTCCGGCTGCAA---CGCGCGCTCGGCTACTCT 291
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 179 AGCGATTCACGAGCTTCGACGCCAAGTCCGCGTCCGAGCGCGCGGTGTCGCTACTCT 238

OY 292 GCGCCGACACAGACTCCCTGGCGGTGAGACAGACATCTCGGCTACGGCTTTCGCGCGACGA 351
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 239 GCGCCGACACAGAGCCCATGGGCTGTGAGACAGACATCTCGGCTACGGCTTTCGCGCGACCT 298

OY 352 GCATCCGCGCGGCGGTCCGAAATCTCTGTGTGCTGCGCGCTCTACGCGCTACCTTCATT 411
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 299 CTATTTCGCGCACCAATAGAGCGCGGTGTGCTGCGCTCTACGAGCTACCTTCATT 358

OY 412 CCGGTCGCGCGCGCGGAGACAAATGTTGTGATCAAGSACACATGGCGCGGACCTG 471
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 359 CCGGTCGCTGTGTGTGAGAAAGATGTTGTGATCAAGTCCACACAGCTGGGCTATCTTG 418

OY 472 GAATTAACAGATTGATATTCGCCATGCCGCGCGCGCTGAGGCAATTCCTAACGCGTGA 531
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 419 GAGAACACCACTTCGATTCATCATATCCCGCGCGCGGTGCGGATCTTCGAGAGTGA 478

OY 532 GCTCGGAGTTGGCGCGCTCCCGCGGCTCAATACGCGGCAATTCGTCGCGGACCACT 591
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 479 CTCCCAAGTGGCGGTCTGCGCGCGCGGCTACGCGCGGCTATCTCTCCGCAACAGT 538

OY 592 GCGATTCTTCCCGCGCGCTCAAGCGCGGTGCGGAGTGGCGGTTGACTGGTCCAGA 651
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 539 GCGATCGGTTCCCGGAGCGCTCAAGCGCGGTGCTACTGAGCGCTTCGACTGGTTCAGA 598

OY 652 AGCGCAACAACCGGAGCTTCAGTCCAGAGTGCAGTCCCGCGCGAGATGTTGCC 711
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 599 AGCGCAACAATCCGAGCTTCAGTCCGTCAGTCCAGTCCGAGCGGAGCTGTGTGCTC 658

OY 712 GCTCGGCGTGCAGAGCGGAGAGAGTCCAGCTTCCCGCTTCACCCCGCAAGCGGTG 771
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 659 GCGCGGAGTGGCGCGCAAGAGAGCAACTTCTCTGCGTCCAGATTCCTCCAGCA 718

OY 772 GCAAGG-----TGCAACGCGGAGCGCGGAGTGGAGTGGCGGTGGT 813
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 719 GCAACAGCTTCGCGGTCAACAGCTTACAGCAACAGCAACAGCTTCACCACT 778

OY 814 CCGGCGAGAC---GTCTCCCGCGCGGAGAGTGGCTGACAGTCTCAGAAATGGGCTAGT 870
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 779 CGAGCCCGCGAGTCCAGCTTACAGTCCGAGCGGTGCACTGTGAGAGTGGGCTAGT 838

OY 871 GCGGTGGAGTCCGCTTCAGAGGATGCAACCACTGTCTGTGGCAGCACTGCGCAAGT 930
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 839 GCGCGCGAGTGGCTGAGAGGCTGACCACTGCGGTGCTGAGCAGCTTTCAGCAAGA 898

OY 931 TGAAGCACTACTACTGCAAGTCCCTTAAAC 961
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 899 TTAAATGATCTGTATCAATCAAGTCCGTAGAC 929
```

```
RESULT 13
US-09-189-060B-55
; Sequence 55; Application US/09189060B
; Patent No. 6270968
; GENERAL INFORMATION:
; APPLICANT: Dalboge, Henrik
; APPLICANT: Sandal, Thomas
; APPLICANT: Kauppinen, Markus
; APPLICANT: Borge, Diderichsen
; TITLE OF INVENTION: Method Of Providing No. 6270968el DNA Sequences
; FILE REFERENCE: 4772.204-US
; CURRENT APPLICATION NUMBER: US/09/189,060B
; PRIOR FILING DATE: 1998-11-10
; PRIOR APPLICATION NUMBER: PCT/DK97/00216
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 55
; LENGTH: 1060
; TYPE: DNA
; ORGANISM: Humicola insolens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (10)...(927)
; NAME/KEY: mat.peptide
; LOCATION: (73)...(927)
; NAME/KEY: sig.peptide
; LOCATION: (10)...(72)
US-09-189-060B-55
```

```
Query Match      37.3%; Score 437.4; DB 4; Length 1060;
Best Local Similarity 70.4%; Pred. No. 3e-78;
Matches 655; Conservative 0; Mismatches 246; Indels 30; Gaps 4;
```

```
OY 55 CCAAGATGCGCTTACTCCGCTTCTTCGCAACACCCGCGCGCTGCACTTCCTGTGCG 114
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5 CCAAGATGCGCTTCTCCCTCCCTCCCTCCGCTGCGCGCTGCGCGCTGCGTGG 64

OY 115 CCTCCGCGCGCAGTGGAGTGGCCAGTCCAGAGATCTGGAGTGTGCAAGCGTCT 174
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 65 CCTTGGCGCGTGTATGGAG-----GTCCACCCGCTACTGTGGAGTGTGCAAGCGTCT 118

OY 175 GCGCTTGGCGCGGAGAGCGCGGTCAAGCAACCGGTCTAGCGGTGGATGCCAATTCC 234
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 119 GCGGCTGGCGGAGAGAGCTCCCTGAGACAGCTGTCTTCTGCGAGCGCAACTTCC 178

OY 235 AGGCGCTGTCCGACTTCATATGTCAGTCCGGCTGCAA---CGCGGCTCGGCTACTCT 291
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 179 AGGCTATCACGAGCTTCGAGCGCAAGTCCGCGTCCGAGCGGCGGTGTCGCTACTCT 238

OY 292 GCGCGGACAGACTCCCTGGCGGTGAGACAGATTCGCTACGCGTTCGCGCGAGCA 351
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 239 GCGCGGACAGACCCCTGGGCTGTGAGAGCAAGCTTCGCGGTGCTGCTGCTGCTCCACT 298

OY 352 GCATTCGCGCGGCGGTCCGAAATCTCTGTGTGCTGCGCGCTCTACGCGCTCACTTCATT 411
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 299 CTATTTCGCGCACCAATAGAGCGCGGTGTGTGCTGCGCTCTACGAGCTCACTTCATT 358

OY 412 CCGGTCGCGCGCGGAGACAAATGTTGTGATCAAGSACACATGGCGCGGACCTG 471
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 359 CCGGTCGCTGTGTGTGAGAAAGATGTTGTGATCAAGTCCACACAGCTGGGCTATCTTG 418

OY 472 GAATTAACAGATTGATATTCGCCATGCCGCGCGGTGAGGAGTGGATTCGAACGCGTGA 531
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 419 GAGAACACCACTTCGATTCATCATATCCCGCGCGGTGCGGATCTTCGAGAGTGA 478

OY 532 GCTCGAGTTCGCGCGCTCCCGGCGCTCAATACGCGGCAATTCCTCCGCGGACCACT 591
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 479 CTCCCAAGTGGCGGTCTGCGCGCGGCTACGCGGCAAGCTGTGTGCTCCGCAACAGT 538

OY 592 GCGATTCTTCCCGCGCGCTCAAGCGCGGTGCGGAGTGGCGGTTGACTGGTCCAGA 651
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```

Db 539 GCGATCGTTCCTCCGAGCGCCCTCAAGCCCGGCTGCTACTGCGCTTCGACGTGTTCAAGA 598
QY 652 AGCGGACAAACCGAGCTTCAGCTTCAGAGAGAGTGCAGTCCCGCCGAGATCGTGGCC 711
Db 599 AGCGGACAAACCGAGCTTCAGCTTCAGAGAGAGTGCAGTCCCGCCGAGATCGTGGCC 658
QY 712 GCTCCGCTGCAAGCGCAACGAGACACTTCCTCCGCTTCACCCCGCCCAAGCGGTG 771
Db 659 GCAACGAGATGCGCCGCAAGGAGGAGCACTTCCTGCGCTTCAGATCCCTTCACAGA 718
QY 772 GCAACG-----TGGCAGCGGAGAGCCGACAGTGCAGTCCGCTGGGT 813
Db 719 GCAACGCTCTCCGCTTCACAGCCTTACAGCACACACACAGCTTCACCTCCACACCT 778
QY 814 CGGAGCAGAGC---GTCTCCCGGCGGCGAGTGCAGTGCAGTTCAGAGTGGGCTCAGT 870
Db 779 CGAGCCCGCCGAGTTCAGAGCTTCAGAGCTCCAGCGGCTGCAGTGCAGAGTGGGCTCAGT 838
QY 871 GCGGTGCAATCGGCTTCAGGAGTGCACACACTGTGTCTGTGGACACCACTTCAGAGAGT 930
Db 839 GCGGCGGCAATGCGTGGAGGCGGTGCACACACTGTGTGTGGAGAGACTTCAGAGAGA 898
QY 931 TGAACGACTACTGCTGCAATGCTGCTTAAC 961
Db 899 TTAATGACTGTACATCACTGCTGTAGAC 929

```

RESULT 14
US-09-230-665-1

```

; Sequence 1, Application US/09230665
; Patent No. 6322595
; GENERAL INFORMATION:
; APPLICANT: Boyer, Stanton L
; TITLE OF INVENTION: Detergent Composition Comprising Two Cellulase
; FILE REFERENCE: 6191 Sequence Listing (8 Sequences)
; Patent No. 6322595
; CURRENT APPLICATION NUMBER: US/09/230,665
; EARLIER FILING DATE: 1999-01-29
; EARLIER APPLICATION NUMBER: 60/023,125
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1060
; TYPE: DNA
; ORGANISM: Humicola insolens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (10)..(924)
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: (73)..(924)
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: (10)..(72)
US-09-230-665-1

```

Query Match 37.3%; Score 437.4; DB 4; Length 1060;
Best Local Similarity 70.4%; Pred. No. 3e-78;
Matches 65; Conservative 0; Mismatches 246; Indels 30; Gaps 4;

```

QY 55 CCAGATGCGCTACTCCGCTTCCTGCAACACCGCGGTGCACTTCCTGTGTCG 114
Db 5 CCAGATGCGCTACTCCGCTTCCTGCAACACCGCGGTGCACTTCCTGTGTCG 64
QY 115 CCTCCGCGGCGAGTGCAGTGGCCAGTCCAGAGATATGAGTGTGCAAGCCGTCT 174
Db 65 CCTCCGCGGCGAGTGCAGTGGCCAGTCCAGAGATATGAGTGTGCAAGCCGTCT 118
QY 175 GCGCTTGGCCCGGAGAGCGCGCTCAGCCAGCGGTCTAGCGGCGATGCCAATTCC 234
Db 119 GCGCTTGGCCCGGAGAGCGCGCTCAGCCAGCGGTCTAGCGGCGATGCCAATTCC 178

```

```

QY 235 AGCGCTGTCCGACTTCAATGTCAGTCCGGCTGCA---CGGCGCTCGGCTACTCT 291
Db 179 AGCTATACAGGACTTCGAGCCCAAGTCCGGCTGCGAGCCGGGCTGTGCTACTCT 238
QY 292 GCGCGGACCAAGACTCCCTGGGGGTGACAGACATCTCGCTACGGCTTGGCGGAGCA 351
Db 239 GCGCGGACCAAGACCCATGGGCTGTACAGCAGACTTGGCTGGCTTGTGCGCACT 298
QY 352 GCATCGCGCGGCGGATCCCTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 411
Db 299 CTATTCGCGGAGCAATATGAGGCGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 358
QY 412 CCGTTCCTCCGCGGCAAGCAATGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 471
Db 359 CCGTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 418
QY 472 GAAATACCAATTCATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 531
Db 419 GAGGCAACCACTTCGATTCACATCCCGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCT 478
QY 532 GCTCGAGTTCGCGGCTCCCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 591
Db 479 CTCCCAAGTTCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 538
QY 592 GCGATTCCTTCCTCCCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 651
Db 539 GCGATGCTTCCTCCCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 598
QY 652 AGCGGACCAACCGAGCTTCAGTTCAGAGTGCAGTCCCGCGGAGATGCTGCTGCC 711
Db 599 AGCGGACCAACCGAGCTTCAGTTCAGAGTGCAGTCCCGCGGAGATGCTGCTGCC 658
QY 712 GCTCCGCTGCAAGCGCAACGAGACACTTCCTGCTGCTGCTGCTGCTGCTGCTGCT 771
Db 659 GCAACGAGATGCGCGCGGCAAGGAGGAGCACTTCCTGCGCTTCAGATCCCTTCAG 718
QY 772 GCAACG-----TGGCAGCGGAGAGCCGACAGTGCAGTCCGCTGGGT 813
Db 719 GACACGCTCTCCGCTGCAACGAGCTTCAGAGCACACACAGCTTCACCTCCACACCT 778
QY 814 GCGGCGAGAGC---GTCTCCCGGCGGCGGAGTGCAGTTCAGTTCAGAGTGGGCTCAGT 870
Db 779 CGAGCCCGCGAGTTCAGAGCTTCAGAGCTTCAGAGCTTCAGAGTGGGCTCAGT 838
QY 871 GCGGTGCAATCGGCTTCAGGAGTGCACACACTGTGTCTGTGGACACCACTTCAGAGAGT 930
Db 839 GCGGCGGCAATGCGTGGAGGCGGTGCACACACTGTGTGTGGAGAGACTTCAGAGAGA 898
QY 931 TGAACGACTACTGCTGCAATGCTGCTTAAC 961
Db 899 TTAATGACTGTACATCACTGCTGTAGAC 929

```

RESULT 15
US-09-189-028-1
; Sequence 1, Application US/09189028
; Patent No. 6423524

```

; GENERAL INFORMATION:
; APPLICANT: Rasmussen, Grethe
; APPLICANT: Mikkelson, Jan Moller
; APPLICANT: Schuelein, Martin
; APPLICANT: Patkar, Shankant A.
; APPLICANT: Hegen, Fred
; TITLE OF INVENTION: A Cellulase Preparation Comprising an
; TITLE OF INVENTION: Endoglucanase Enzyme
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESSES:
; ADDRESS: No. 64235240 No. 6423524disk of No. 6423524th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America

```

```

: ZIP: 10174-6401
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/189,028
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/389,423
: FILING DATE: 14-FEB-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Lambiris, Elias J.
: REGISTRATION NUMBER: 33,728
: REFERENCE/DOCKET NUMBER: 3469, 214-US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 212-867-0123
: TELEFAX: 212-878-9655
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1060 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: HYPOTHEetical: NO
: ORIGINAL SOURCE:
: ORGANISM: Humicola insolens
: STRAIN: DSM 1800
: FEATURE:
: NAME/KEY: mat-peptide
: LOCATION: 73..924
: FEATURE:
: NAME/KEY: s19-peptide
: LOCATION: 10..72
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 10..924
: US-09-189-028-1

```

```

Query Match      37.3%; Score 437.4; DB 4; Length 1060;
Best Local Similarity 70.4%; Pred. No. 3e-78;
Matches 655; Conservative 0; Mismatches 246; Indels 30; Gaps 4;

```

```

QY 55 CCAAGATGCGCTCTACCTCCGCTTCTTGCCACAAACCCCTGCGCCTGCTGCTCTGCTG 114
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5 CCAAGATGCGCTCTCTCCCTCCCTCTCCGCTGCGCGCTTGCGCCCTGCGCGCTGTTGG 64

QY 115 CCTCCGCGGCGCAGTGGCAGTGCACAGATGCTGGAAGTCTGCTGCAAGCGCTGCT 174
   || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 65 CCTTGGCGCGTGAATGGAG-----GTCCACCGCGTACTGGGACTGCTGCAAGCTTCTG 118

QY 175 GCGCTTGGCCCGGAGAGCGCGCGCTCAGCCACCGGCTTAAGCGGTGGATGCCAATTTC 234
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 119 GCGGCTGGGCGCAAGAGGCTCCCGTGAACCAAGCCTGCTTTCTTCCGAACGCCAATTCC 178

QY 235 AGGCGCTGTCCGACTTCAATGTCCAGTGGGCTGCA-----CGGCGGCTCGGCTACTTCT 291
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 179 AGCGTATCAAGGACTTCAAGTCCGCGTGGCGAGCCGGGCGGTGCTGCGCTTACTGCT 238

QY 292 GCGCGCAGCAGACTCCCTGCGGCGGTGAGACAACTCTGCGCTACGGCTTGGCGCGAGA 351
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 239 GCGCGCAGCAGACCCCATGGGCTGTGAAGACGACTTCCGCGTGGGTTTGTGCGCACCT 298

QY 352 GCATCGCGCGGCGGTCCGATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 411
   || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 299 CTAATGCGCGCAGCAATGAGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 358

QY 412 CCGGTCCTGCTGCGCGCAGCAAGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 471
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 359 CCGGTCCTGCTGCTGCGCAAGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 418

```

```

QY 472 GAATTAACCACTTGCATATCCCATGCCCGCGCGCGCTGGGCAATCTTCAAGCGCTGCA 531
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 419 GCAGCAACCACTTGCATCTCAATCCCCGGCGCGCGCTGGGCAATCTTCAAGCGATGCA 478

QY 532 GCTGCAGTTCGGCGCGCGCTCCCGCGCGCTCAATAGGGGGGCAATTTCGTCGCCGCAAGCT 591
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 479 CTCGCCAATTTGGCGGCTTGGCCCGCGCGCTGAGCGCTGAGGGGCAATCTTCCGCAAGG 538

QY 592 GCGATTTCTTCCCGCGCGCGCTCAAGCCCGCGCTGAGTGGCGGTTTGATGTTCCAGA 651
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 539 GCGATCGGTTCCCGCGCGCGCTCAAGCCCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 598

QY 652 ACGCGCAACCCGAGCTTCAAGCTTCAAGAGTGCAGTGCCTCCCGCGCGAGATCGTTGCC 711
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 599 ACGCGCAACCTCCAGCTTCAAGCTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 658

QY 712 GCTCCGCGCTGCAAGCGCAAGCAAGCTCCAGTTCCTCCGCTTCAACCCCGCAAGCGGTG 771
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 659 GCACCGGATGCGCGCGCGCGCAACGACGCAAGCTTCCCTGCTGCTGCTGCTGCTGCTG 718

QY 772 GCAACGG-----TGGCACCGGAGCGCCACGCTGCACTGCGCGCTGGGT 813
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 719 GCACCAAGCTTCCGCTCAACCAAGCTTACCAAGCAAGCAAGCAAGCAAGCTTCCACACT 778

QY 814 CGGCGCAAGAC---GTCTCCCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 870
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 779 CGAGCGCGCGCTCAGCTTCAAGCTTCCAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 838

QY 871 GCGGTGCGATGCGCTTCAAGCGATGCAACCACTGCTGCTGCTGCTGCTGCTGCTGCTG 930
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 839 GCGGCGGCAATGCTGAGGCGGCTGCAACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 898

QY 931 TGAACGACTTACTGCGAGTGCCTTAAC 961
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 899 TTAATGACTGTACCAATCAGTGCCTGTAGAC 929

```

Search completed: February 27, 2003, 17:47:34
Job time : 59 secs

•
•
•
•

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2003, 17:47:40 ; Search time 3182 Seconds

(Without alignments)
10737.480 Million cell updates/sec

Title: US-10-007-521-11

Perfect score: 1174
Sequence: 1 GAGCAGCACCCTCAAGCTG.....AAAAAAAAAAAAAAAAAAAA 1174

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 segs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: *
1: gb_da: *
2: gb_htg: *
3: gb_in: *
4: gb_om: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vl: *
15: em_da: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_mu: *
20: em_om: *
21: em_or: *
22: em_ov: *
23: em_pat: *
24: em_ph: *
25: em_pl: *
26: em_ro: *
27: em_sts: *
28: em_un: *
29: em_vl: *
30: em_htg_hum: *
31: em_htg_inv: *
32: em_htg_other: *
33: em_htg_mus: *
34: em_htg_pln: *
35: em_htg_rod: *
36: em_htg_mam: *
37: em_htg_vrt: *
38: em_sy: *
39: em_higo_hum: *
40: em_higo_mus: *
41: em_higo_other: *

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1174	100.0	1174	A92311	A92311 Sequence 1
2	1174	100.0	1174	AR075389	AR075389 Sequence
3	1174	100.0	1174	AR094310	AR094310 Sequence
4	451.6	38.5	922	A68070	A68070 Sequence 65
5	451.6	38.5	922	AR163167	AR163167 Sequence
6	437.4	37.3	1060	A21793	A21793 H. insolens
7	437.4	37.3	1060	A23635	A23635 H. insolens
8	437.4	37.3	1060	A23644	A23644 H. insolens
9	437.4	37.3	1060	A23953	A23953 Endoglucanase
10	437.4	37.3	1060	A23957	A23957 Endoglucanase
11	437.4	37.3	1060	A41658	A41658 Sequence 1
12	437.4	37.3	1060	A68060	A68060 Sequence 55
13	437.4	37.3	1060	AR059002	AR059002 Sequence
14	437.4	37.3	1060	AR072921	AR072921 Sequence
15	437.4	37.3	1060	AR163160	AR163160 Sequence
16	437.4	37.3	1060	BD010852	BD010852 Cellulase
17	437.4	37.3	1060	I13884	I13884 Sequence 1
18	437.4	37.3	1060	I21316	I21316 Sequence 1
19	437.4	37.3	1060	I57983	I57983 Sequence 1
20	428.6	36.5	1057	BD002248	BD002248 Cellulase
21	411.4	35.0	919	A68072	A68072 Sequence 67
22	411.4	35.0	919	AR163168	AR163168 Sequence
23	392	33.4	922	A68078	A68078 Sequence 73
24	392	33.4	922	AR163171	AR163171 Sequence
25	377.2	32.1	1257	AR121354	AR121354 Sequence
26	374.6	31.9	1904	AB003107	AB003107 Humicola
27	369.8	31.5	1154	AR094308	AR094308 Sequence
28	364.6	31.1	924	A68074	A68074 Sequence 69
29	364.6	31.1	924	AR163169	AR163169 Sequence
30	337	28.7	894	AR094306	AR094306 Sequence
31	336.2	28.6	912	A68076	A68076 Sequence 71
32	336.2	28.6	912	AR163170	AR163170 Sequence
33	334.8	28.5	927	AR094307	AR094307 Sequence
34	315.2	26.8	1443	FSOKCH	L29381 Fusarium ox
35	315.2	26.8	1473	A21795	A21795 F. oxysporum
36	315.2	26.8	1473	A23637	A23637 F. oxysporu
37	315.2	26.8	1473	A23646	A23646 F. oxysporu
38	315.2	26.8	1473	A23955	A23955 Endoglucanase
39	315.2	26.8	1473	A23959	A23959 Endoglucanase
40	315.2	26.8	1473	A41660	A41660 Sequence 3
41	315.2	26.8	1473	AR072922	AR072922 Sequence
42	315.2	26.8	1473	BD002249	BD002249 Cellulase
43	315.2	26.8	1473	BD010853	BD010853 Cellulase
44	315.2	26.8	1473	I13885	I13885 Sequence 3
45	315.2	26.8	1473	I21317	I21317 Sequence 3

ALIGNMENTS

RESULT 1
LOCUS A92311 1174 bp DNA linear PAT 22-JAN-2000
DEFINITION Sequence 1 from Patent EP0843041.
ACCESSION A92311
VERSION A92311.1 GI:6741083
KEYWORDS
SOURCE
ORGANISM baker's yeast.
Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE
1 (bases 1 to 1174)
AUTHORS Kallum, L. and Lund, H.
TITLE
Garments with considerable variation in abrasion level and process
for its production using cellulolytic enzymes

Oy 421 TCGCCGCAAGCAATGTTGTCAGTCAACGAGCACTGGCGGCACTGGAGTAAC 480
 |||||||
 Db 421 TCGCCGCAAGCAATGTTGTCAGTCAACGAGCACTGGCGGCACTGGAGTAAC 480
 |||||||
 Oy 481 AGTTCGATATGTCATGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 540
 |||||||
 Db 481 AGTTCGATATGTCATGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 540
 |||||||
 Oy 541 TCGCGGCG 600
 |||||||
 Db 541 TCGCGGCG 600
 |||||||
 Oy 601 TCGCGGCG 660
 |||||||
 Db 601 TCGCGGCG 660
 |||||||
 Oy 661 ACCGAGCTTACGTTCCAGCAGGTCAGTCCCGCGCGAATCGTTCCCGCGCGCT 720
 |||||||
 Db 661 ACCGAGCTTACGTTCCAGCAGGTCAGTCCCGCGCGAATCGTTCCCGCGCGCT 720
 |||||||
 Oy 721 GCAAGCGCAAGCACTCCAGTTCCTCCGCTTCACCGCGCGCGCGCGCGCGCG 780
 |||||||
 Db 721 GCAAGCGCAAGCACTCCAGTTCCTCCGCTTCACCGCGCGCGCGCGCGCGCG 780
 |||||||
 Oy 781 GCACCGGAGCG 840
 |||||||
 Db 781 GCACCGGAGCG 840
 |||||||
 Oy 841 GTGCTGTCAGTTCAGAAAGTGGCTCAGTGGCGCGCGCGCGCGCGCGCGCG 900
 |||||||
 Db 841 GTGCTGTCAGTTCAGAAAGTGGCTCAGTGGCGCGCGCGCGCGCGCGCGCG 900
 |||||||
 Oy 901 CCGTGTCTCTGGCAGCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 960
 |||||||
 Db 901 CCGTGTCTCTGGCAGCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 960
 |||||||
 Oy 961 CAGTTTTCGACAGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1020
 |||||||
 Db 961 CAGTTTTCGACAGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1020
 |||||||
 Oy 1021 TTTTTCGAGCGCTCAATACATACATACCTTCGATTCGATACAGCGCGTAC 1080
 |||||||
 Db 1021 TTTTTCGAGCGCTCAATACATACATACCTTCGATTCGATACAGCGCGTAC 1080
 |||||||
 Oy 1081 TCTCAGCAGCAGTTCGGGCGCGAGTCAAGCCGTTTAAACGCGCGCGCG 1140
 |||||||
 Db 1081 TCTCAGCAGCAGTTCGGGCGCGAGTCAAGCCGTTTAAACGCGCGCGCG 1140
 |||||||
 Oy 1141 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1174
 |||||||
 Db 1141 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1174
 |||||||

RESULT 3
 AR094310
 LOCUS AR094310 1174 bp DNA linear PAT 08-SEP-2000
 DEFINITION Sequence 11 from patent US 6001639.
 ACCESSION AR094310 GI:10021119
 VERSION AR094310.1
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE
 AUTHORS 1 (bases 1 to 1174)
 Schluep, M., Andersen, L., Nonboe, J., Lassen, S., Slashedren, F.,
 Kauppinen, M., Sakari, L., Lange, L., Nielsen, R., Ilum, J., Thara, M. and
 Takagi, S.
 TITLE Endoglucanases
 JOURNAL Patent: US 6001639-A 11 14-DEC-1999;
 FEATURES
 source 1..1174
 location/Qualifiers
 BASE COUNT 243 a 395 c 320 g 216 t

ORIGIN

Query Match 100.0%; Score 1174; DB 6; Length 1174;
 Best Local Similarity 100.0%; Pred. No. 7,56-197;
 Matches 1174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GAGCAGCAACCCCTCAAGCTGTCAGTTCACACCCGCTCTCTTTCTTGGCCCCAGCA 60
 |||||||
 Db 1 GAGCAGCAACCCCTCAAGCTGTCAGTTCACACCCGCTCTCTTTCTTGGCCCCAGCA 60
 |||||||
 Oy 61 TCGCTCTACTCCGTTTCTTCACACACCTGGCGCGCTCACTTCCTGCGCGCTCG 120
 |||||||
 Db 61 TCGCTCTACTCCGTTTCTTCACACACCTGGCGCGCTCACTTCCTGCGCGCTCG 120
 |||||||
 Oy 121 CGGCGAGTGCAGTGGCGAGTCCACAGATATCGGAGTGTGCAAGCGCTGTGCGCTT 180
 |||||||
 Db 121 CGGCGAGTGCAGTGGCGAGTCCACAGATATCGGAGTGTGCAAGCGCTGTGCGCTT 180
 |||||||
 Oy 181 GCGCGGGAAGCGCGCTCAAGCCAGCGTACGCTGGGATGCAACTTCACAGCGC 240
 |||||||
 Db 181 GCGCGGGAAGCGCGCTCAAGCCAGCGTACGCTGGGATGCAACTTCACAGCGC 240
 |||||||
 Oy 241 TGTCCGACTTCATATGTCAGTCCGCGCTCAAGCGCGCTCGGCTCTCTGCGCGAC 300
 |||||||
 Db 241 TGTCCGACTTCATATGTCAGTCCGCGCTCAAGCGCGCTCGGCTCTCTGCGCGAC 300
 |||||||
 Oy 301 AGACTCCCTGGCGGTGACAGCAATCTGCTTACGCGCTTGGCGGAGCAATCGCG 360
 |||||||
 Db 301 AGACTCCCTGGCGGTGACAGCAATCTGCTTACGCGCTTGGCGGAGCAATCGCG 360
 |||||||
 Oy 361 GCGGATCGCAATTCGCGGTGCGCGCTGCTACGCGCTCACTTCCTCGGCTCGCG 420
 |||||||
 Db 361 GCGGATCGCAATTCGCGGTGCGCGCTGCTACGCGCTCACTTCCTCGGCTCGCG 420
 |||||||
 Oy 421 TCGCGGCAAGCAATGTTGTCAGTCAACGAGCACTGGCGGAGCTGGAGTAAC 480
 |||||||
 Db 421 TCGCGGCAAGCAATGTTGTCAGTCAACGAGCACTGGCGGAGCTGGAGTAAC 480
 |||||||
 Oy 481 AGTTCGATATGTCATGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 540
 |||||||
 Db 481 AGTTCGATATGTCATGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 540
 |||||||
 Oy 541 TCGCGGCGCTCCCGCGCGCTCAATACGCGCGCTTTCGCGGAGCGCATTCCT 600
 |||||||
 Db 541 TCGCGGCGCTCCCGCGCGCTCAATACGCGCGCTTTCGCGGAGCGCATTCCT 600
 |||||||
 Oy 601 TCGCGGCGCTCCCGCGCGCTCAATACGCGCGCTTTCGCGGAGCGCATTCCT 660
 |||||||
 Db 601 TCGCGGCGCTCCCGCGCGCTCAATACGCGCGCTTTCGCGGAGCGCATTCCT 660
 |||||||
 Oy 661 ACCGAGCTTACGTTCCAGCAGGTCAGTCCCGCGCGAATCGTTCCCGCGCGCT 720
 |||||||
 Db 661 ACCGAGCTTACGTTCCAGCAGGTCAGTCCCGCGCGAATCGTTCCCGCGCGCT 720
 |||||||
 Oy 721 GCAAGCGCAAGCACTCCAGTTCCTCCGCTTCACCGCGCGCGCGCGCGCGCG 780
 |||||||
 Db 721 GCAAGCGCAAGCACTCCAGTTCCTCCGCTTCACCGCGCGCGCGCGCGCGCG 780
 |||||||
 Oy 781 GCACCGGAGCG 840
 |||||||
 Db 781 GCACCGGAGCG 840
 |||||||
 Oy 841 GTGCTGTCAGTTCAGAAAGTGGCTCAGTGGCGCGCGCGCGCGCGCGCGCG 900
 |||||||
 Db 841 GTGCTGTCAGTTCAGAAAGTGGCTCAGTGGCGCGCGCGCGCGCGCGCGCG 900
 |||||||
 Oy 901 CCGTGTCTCTGGCAGCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 960
 |||||||
 Db 901 CCGTGTCTCTGGCAGCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 960
 |||||||
 Oy 961 CAGTTTTCGACAGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1020
 |||||||
 Db 961 CAGTTTTCGACAGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1020
 |||||||

QY	1021	TTTTGGGGCGCTCAATCATACATCAATCAACCTTCGATTCCTTGACATATAGCAGCCGGTACACA	1080
Db	1021	TTTTTGGAGCGCTCAATCATACATCAATCAACCTTCGATTCCTTGACATATAGCAGCCGGTACACA	1080
QY	1081	TCTCAGACCGACTTTGGGGGGGGAATCAGCGCCGCTTTTAAAAA	1140
Db	1081	TCTCAGACCGACTTTGGGGGGGGAATCAGCGCCGCTTTTAAAAA	1140
QY	1141	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	1174
Db	1141	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	1174
RESULT 4			
LOCUS	A68070	922 bp	DNA
DEFINITION	Sequence 65 from Patent WO9743409.	linear	PAT 05-MAY-1999
ACCESSION	A68070		
VERSION	A68070.1	GI:4756870	
KEYWORDS			
SOURCE	unidentified.		
ORGANISM	unclassified.		
REFERENCE	1 (bases 1 to 922)		
AUTHORS	Dalboeuf, H., Diderichsen, B., Sandal, T. and Kauppinen, S.		
TITLE	METHOD OF PROVIDING NOVEL DNA SEQUENCES		
JOURNAL	Patent: WO 9743409-A 65 20-NOV-1997;		
FEATURES	NOVONORDISK AS (DK)		
source	Location/Qualifiers		
	1..922		
	/organism="unidentified"		
	/db_xref="taxon:32644"		
	<1..922		
	/note="unnamed protein product"		
	/codon_start=2		
	/protein_id="CAB42308.1"		
	/db_xref="GI:4756871"		
	/translation="PFMMVAVMMSLFYGLGVAAAPAFADGRSTRYMDCKRPSGWPGR		
	ALYNQPYVARNANFORITDDPKASGCGGSGASCADOTPMVAVSDPFAVGAATALAGQ		
	SESWCCACAELEFTTSGPVGAKKHAQVOSTSGDGLGNHFDLNNPFGGVGIFPGCSPO		
	VGGIAGRTGGSVSRSEDSDFPALAKTCTRYRDMFKNADNPFSFQVOCQPELVAR		
	TGCRNDGNEPVAQIIPSSSTSPSPVNPQCTSTSTSTSTSTSPVQPTTSGCTAERMA		
	QCGNGMSGCTTCVAGSGCTCKINDMYHQL"		
BASE COUNT	151 a 333 c 262 g 176 t		
ORIGIN			
Query Match	38.5%	Score 451.6:	DB 6; Length 922;
Best Local Similarity	71.6%;	Pred. No. 1e-69;	
Matches 630: Conservative	0;	Mismatches 229;	Indels 21; Gaps 2;
QY	102	CTTCTCTGATGATGCGCTCCGCGGCGAGTGGCCAGTGGCCAGTCCACGAGATCTAGGAGCTGC	161
Db	43	CTTCAGATGGGGCAACGTCTTCTGCTGTATGGCAGAGTCCACCGGACTGGGATTTG	102
QY	162	TCCAAAGCTGCTGGCTTTGGCCCGGGAAGCCGCCGCTCAGCCACACCGCTACAGCGTGC	221
Db	103	TGTAAAGCCGTCGTGTCGGCGCCGGAAGGGGCTGTGAACCAACCGCTTACGCCCGC	162
QY	222	GATGCAACTTCCAGAGCGCTGTCCGACTTCATATGTCACAGTCGGGCTGCMAACGGCGCTCG	281
Db	163	AACGCAACTTCCAGAGGATTCACCGACCCCAACGCGCAAGTCCGGCTGGGATGGCGCTCC	222
QY	282	GCTACTCTCGGCGCGACCAAGACTCCCTTGGCGGGGTGAACAGCAATTCGCTTACGGCTTC	341
Db	223	GCTTCTCTCGGCGCGACCAAGACCCCGTGGCGGTGAAGAGACTTTGCCATACGGTTTC	282
QY	342	GGCGGAGAGAGATGGCGGGGGGTCCGAATCCTGTGTGTGCTGGCGCCGCTACAGCGCTC	401
Db	283	GGGCGTACGGCGCTGCGCGGCAAGTCCGAGTCTTGTGCTGTGCTGCTGTACGAACTC	342
QY	402	ACCTTCACTTCGCGTCCGCTCGCGGCAAGACATGCTGTCGATGCAACGACACTGGC	461

Db	343	ACCTTCACTTGGGGCCCCGTTGCTGGGCAAGAAGATGGCTGTCAGTCCACAGTCCACAGCAGCATGGC	402
OY	462	GGCGACCTGGGAAGTAACCAAGTTGATATGCGCAATGCCCGCGCGGGGGGAGATCTTC	521
Db	403	GGTGAAGCTGGGTAGCAACCACTTTGACCTCAACATGGCCAGATGGCGGGTGGGATCTTC	462
OY	522	AAGCGCTGCACCTGCGCAGTTGGGGGCGCTCCCGGCGCTCAATAGGCGCATTTTCGTG	581
Db	463	GACGGCTGTGCGCTCAGAGTTGGCGGTTCCTCCGCGCAGCGCTATGGGGGGTCTCGCC	522
OY	582	CGCGACAGTGCATTCCTTCCCGCGCGCTCAAGCCCGGCTGCCAGTGGCGGTTTGAC	641
Db	523	CGCAGCGAATGCGAATCCTTCCCGCGGCACTCAAGCCCGGCTGCTACTGCGCTACGAC	582
OY	642	TGGTTCCAGAAACCGCGACAACCCGACGTTCAAGTTCCAGCAGGTGCAATGCCCGCGCAG	701
Db	583	TGGTTTAAAGACGCGCGAATCCGAGCTTTCAGCTTCCCTCAGGTGCATGTGCCACGCGAG	642
OY	702	ATCGTTGCCCCGCTCGGGCTGCAGAGCGCAACGACGTCAGTTGCCCGGTCTTCAACCCC	761
Db	643	CTCGTCGCTCGGACCGGATGCGCGCGCAACGACGAGCGGCACTTCTCGCTCGCATATC	702
OY	762	CCAAACGGGGAACGCG-----TGGCACCGGGAGCGCCACGTCGACT	803
Db	703	CCCTCAGACGACCAACGACTCTCCGGTCAACACGAGCTACAGAGCAACGACACGATCCACC	762
OY	804	GGCGCTGGGTCGGGCGAAGC--GTCTCCCGGGCGGCGAATGGCTGCAGCTCTAGAG	860
Db	763	TCCACACACTCGAGCCCGCAGTCCAGGCTTACGACTCCAGGCGGTGCGACATGCTTAGAGG	822
OY	861	TGGGCTCAGTGGGTTGGCATTCGGCTTCAGCGGAGATCACACCTGTGTCMTGGGACGACC	920
Db	823	TGGGCTCAGTGGGCGGCGAATGGCTGGAGACGGGTGCACACCTGGGTGCTGTGGCAGACT	882
OY	921	TGCCAGAGATTGAAGCACTACTCTCGCATGCTCTAAA	960
Db	883	TGCACGAGATTATATGACTGTATCATCATGATGCTCTGTANA	922

LOCUS	ARI63167	922 bp	DNA	Linear	PAT 17-OCT-2001
DEFINITION	Sequence	65	from patent US 6270968.		
ACCESSION	ARI63167				
VERSION	ARI63167.1	GI:16233670			
KEYWORDS					
SOURCE	Unknown.				
ORGANISM	Unknown.				
REFERENCE	Unclassified.				
AUTHORS	1 (bases 1 to 922) Dalb.o slashed.ge,H., Sandal,T., Kauppinen,M.Sekari, and Dierichsen,Balashedge.				
TITLE	Method of providing a hybrid polypeptide exhibiting an activity of interest				
JOURNAL	Patent: US 6270968-A 65 07-AUG-2001;				
FEATURES	Location/Qualifiers				
source	1..922				
BASE COUNT	151 a 333 c 262 g 176 t				
ORIGIN	/organism="unknown"				
Query Match	38.5%;	Score 451.6;	DB 6;	Length 922;	
Best Local Similarity	71.6%;	Pred. No.1e-69;			
Matches 630;	Conservative 0;	Mismatches 229;	Indels 21;	Gaps 2	
QY 102	CTTCCCTGCTGGTGGCCGCGGCGGACGAGGAGCGGACGACGACGAGATCTGGAGCTGC	161			
DB 43	CTTCAAGTCCGGGACACCTGCTTTCGCTCGTGAAGGACAGGCTCCAGCGGATAGTGGATTGG	102			
QY 162	TGCAAGCGCTGTCGCGCTGGTGGCCCGGAGGAGCGCCGCTGACGACCGGCTACGCGTGC	221			
DB 103	TGTAAAGCGCTGTCGTCGTGTCGCGGAGCGGCGTGTGTAACGACCGCGCTACGCGCGC	162			

QY 931 TGAACGACTACTGCGAGGCGCTTAAC 961
 Db 899 TTAATGACTGTTACATCACTGCTGTAGAC 929
 RESULT 7
 LOCUS A23635 1060 bp mRNA linear PAT 23-FEB-1995
 DEFINITION H. insolens endoglucanase gene.
 ACCESSION A23635.1 GI:832886
 VERSION A23635.1 GI:832886
 KEYWORDS Humicola insolens.
 SOURCE Humicola insolens.
 ORGANISM Humicola insolens.
 Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Humicola.
 REFERENCE 1 (bases 1 to 1060)
 AUTHORS Convents, A.C., Busch, A. and Baeck, A.C.
 TITLE Detergent compositions with high activity cellulase and softening
 clays
 JOURNAL Patent: EP 0495258-A 1 22-JUL-1992;
 THE PROCTER & GAMBLE COMPANY
 FEATURES
 source location/Qualifiers
 1..1060
 /organism="Humicola insolens"
 /db_xref="taxon:34413"
 10..927
 /codon_start=1
 /product="endoglucanase"
 /protein_id="CAA01681.1"
 /db_xref="GI:832887"
 CDS
 55 CGAGATGCGCTACGCTCCGCTTTCGACACACCGCGCGCTGCTCTCTGCTG 114
 Db 5 CCAAGATGCGCTTCTCCCTCCCTCCCTGCGCGCTGCTGCGCGCTGCTGCG 64
 QY 115 CTTCCGCGCGCGAGTGGCGAGTGCACGAGATGCTGAGTGTGCAAGCCGTGT 174
 Db 65 CCCTTGGCGCTGATGGAG-----GTCCACCGCGCTACTGGAGTGTGCAAGCCGTGT 118
 QY 175 GCGGTGGCGCGGGAAGGCGCGCTGACCAACCGGTCTACCGCTGCATGGCACTTCC 234
 Db 119 GCGGCTGGCGCAAAAGCGCTCCGTGAACACGCTGCTTTCGCAACGCCAATTTC 178
 QY 235 AGCGCTGCTCGCATGCTCAATGTCCAGTGGGCTGCA--CGGCGGCTCGGCTACTCT 291
 Db 179 AGCGTATCAGGAGTTCGACGCCCAAGTCCGGCTGAGACCGCGCGGTGTGCTACTGCT 238
 QY 292 GCGCGGACCAAGTCTCCGCGGTGGAAGCAATCTCGCTACGGCTTCGCCGAGCA 351
 Db 239 GCGCGGACCAAGTCTCCGCGGTGGAAGCAATCTCGCTACGGCTTCGCCGAGCA 298
 QY 352 GCAATGCGCGGCGGAGTCTGCTGCTGCTGCGCTGCTGCTGCTGCTGCTGCTGCT 411
 Db 299 CTATTGCGCGGAGCAATGAGCGGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 358
 QY 412 CCGGTCCGCTGCGCGGAGCAATGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 471
 Db 359 CCGGTCTGCTGCTGCGGAGCAATGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTG 418
 QY 472 GAACTAACCAATGCTGATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 531

Db 419 GAGCAACCACTGATGATTCATCAATCCCGCGCGCGCTGTGCTGCTGCTGCTGCTG 478
 QY 532 GCTCGAGTTCGCGCGCGCTCCCGCGCGCTCAATAGCGCGCGCTTTCGCGCGAGCAG 591
 Db 479 CTCCCGAGTTCGCGCGCGCTCCCGCGCGCTCAATAGCGCGCGCTTTCGCGCGAGCAG 538
 QY 592 GCGATTCCTTCGCGCGCGCTCCCGCGCGCTCAATAGCGCGCGCTTTCGCGCGAGCAG 651
 Db 539 GCGATTCGCTTCGCGCGCGCTCCCGCGCGCTCAATAGCGCGCGCTTTCGCGCGAGCAG 598
 QY 652 AGCGCGCAACCCGCGCGCTTCAGTTCAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 711
 Db 599 AGCGCGCAATCCGCGCGCTTCAGTTCAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 658
 QY 712 GCTCGCGCGCAAGCGCGAGTTCAGTTCAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 771
 Db 659 GCAACCGGATGCGCGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCG 718
 QY 772 GCAACG-----TGCAACCGGAGCGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCG 813
 Db 719 GCAACGCTGCTCCGCTCAACGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAG 778
 QY 814 CCGCGCGAGAC---GTCTCCCGCGCGCGCGAGTGTGCTGCTGCTGCTGCTGCTGCTG 870
 Db 779 CGAGCGCGCGAGTGCACGCGCTGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAG 838
 QY 871 GCGGTGCGATGCGCTTCAGCGAGTGCACGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAG 930
 Db 839 GCGCGCGCAATGCTGCTGAGCGCGCTGCAACGAGCGAGCGAGCGAGCGAGCGAGCGAG 898
 QY 931 TGAACGACTACTGCGAGTGCCTCTTAAC 961
 Db 899 TTAATGACTGTTACATCACTGCTGTAGAC 929
 RESULT 8
 LOCUS A23644 1060 bp mRNA linear PAT 23-FEB-1995
 DEFINITION H. insolens endoglucanase gene.
 ACCESSION A23644
 VERSION A23644.1 GI:832890
 KEYWORDS Humicola insolens.
 SOURCE Humicola insolens.
 ORGANISM Humicola insolens.
 Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Humicola.
 REFERENCE 1 (bases 1 to 1060)
 AUTHORS Convents, A.C., Busch, A. and Baeck, A.C.
 TITLE Detergent compositions with high activity cellulase and quaternary
 ammonium compounds
 JOURNAL Patent: EP 0495554-A 1 22-JUL-1992;
 THE PROCTER & GAMBLE COMPANY
 FEATURES
 source location/Qualifiers
 1..1060
 /organism="Humicola insolens"
 /strain="DSM 1800"
 /db_xref="taxon:34413"
 10..927
 /codon_start=1
 /product="endoglucanase"
 /protein_id="CAA01683.1"
 /db_xref="GI:832891"
 CDS
 55 CGAGATGCGCTACGCTCCGCTTTCGACACACCGCGCGCTGCTCTCTGCTG 114
 Db 5 CCAAGATGCGCTTCTCCCTCCCTCCCTGCGCGCTGCTGCGCGCTGCTGCG 64
 QY 115 CTTCCGCGCGCGAGTGGCGAGTGCACGAGATGCTGAGTGTGCAAGCCGTGT 174
 Db 65 CCCTTGGCGCTGATGGAG-----GTCCACCGCGCTACTGGAGTGTGCAAGCCGTGT 118
 QY 175 GCGGTGGCGCGGGAAGGCGCGCTGACCAACCGGTCTACCGCTGCATGGCACTTCC 234
 Db 119 GCGGCTGGCGCAAAAGCGCTCCGTGAACACGCTGCTTTCGCAACGCCAATTTC 178
 QY 235 AGCGCTGCTCGCATGCTCAATGTCCAGTGGGCTGCA--CGGCGGCTCGGCTACTCT 291
 Db 179 AGCGTATCAGGAGTTCGACGCCCAAGTCCGGCTGAGACCGCGCGGTGTGCTACTGCT 238
 QY 292 GCGCGGACCAAGTCTCCGCGGTGGAAGCAATCTCGCTACGGCTTCGCCGAGCA 351
 Db 239 GCGCGGACCAAGTCTCCGCGGTGGAAGCAATCTCGCTACGGCTTCGCCGAGCA 298
 QY 352 GCAATGCGCGGCGGAGTCTGCTGCTGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 411
 Db 299 CTATTGCGCGGAGCAATGAGCGGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 358
 QY 412 CCGGTCCGCTGCGCGGAGCAATGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 471
 Db 359 CCGGTCTGCTGCTGCGGAGCAATGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTG 418
 QY 472 GAACTAACCAATGCTGATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 531

Query Match 37.3%; Score 437.4; DB 6; Length 1060;
 sig_peptide 190 a 377 c 288 g 205 t
 BASE COUNT 190 a 377 c 288 g 205 t
 ORIGIN

Db 599 ACGCCGACATCCGAGCTTCAGCTTCGTCAGGTCCAGTCCACGCGAGCTGCTGCTC 658
 Oy 712 GCTCCGGCTGCAAGGCGCAAGAGCTCCAGCTTCCTCCGCTTCCTACCCGCCCAAGCGGTG 771
 Db 659 GCACCGGAGTCCGGCGCGCAAGAGAGCTTCCTCCGCTTCGCTCCAGATCCCTCCAGCA 718
 Oy 772 GCAACGG-----TGGCAGCGGGAGCGCCACAGTGCATGCGCTGGGT 813
 Db 719 GCACACGCTCTCCGGTCAACACAGCTTACAGCAGACAGCTGCTCCAGCACT 778
 Oy 814 GGGGCGAGC-----GTCTCCGGGGGGGGGAGTGGCTGCTGCTCAGAGAGGCTCAGT 870
 Db 779 CGAGCCCGCAGTCCAGCTCAGCTCCAGCGGCTGCTACTGCTGAGAGAGGCTCAGT 838
 Oy 871 GCGGTGCTATCGCTTCAGCGGATGACACACCTGTGTCTGTGCGACCACTCCAGAGT 930
 Db 839 GGGGCGGCAATGGCTGAGGGGCTGCACACCTGCGCTGCGCAGCACTTGCACGAAGA 898
 Oy 931 TGAACGACTACTGCTGCAAGTGCCTTAAC 961
 Db 899 TTAATGACTGTACCAATGACTGCTGTAGAC 929

RESULT 10
 A23957
 LOCUS A23957 1060 bp DNA linear PAT 01-MAR-1995
 DEFINITION Endoglucanase gene.
 ACCESSION A23957
 VERSION A23957.1 GI:832898
 KEYWORDS
 SOURCE Humicola insolens.
 ORGANISM Humicola insolens.
 Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Humicola.
 REFERENCE 1 (bases 1 to 1060)
 AUTHORS McCordudale, F., and Busch, A.
 TITLE Dye transfer inhibiting compositions
 JOURNAL Patent: EP 0540784-A 1 12-MAY-1993;
 THE PROCTER & GAMBLE COMPANY
 FEATURES
 source
 location/Qualifiers
 1..1060
 /organism="Humicola insolens"
 /db_xref="taxon:34413"
 10..927
 /codon_start=1
 /product="endoglucanase"
 /protein_id="CA01698.1"
 /db_xref="GI:832898"
 /translation="MRSSPLPSAVVAALPVLALADGRSTRYWDCKPSCGAKKAP
 VNQPVSCNANFORITDFDAKSGCEPGVAVSCADQTPMVAVNDFAFGAFATSIAGSN
 EAGWCACYELEFTSGPVAGKKNVOSTSTGSDLSNHFIDINIGGAGVIGDCTPQF
 GCIPIGRYGGIISRNEDDRPDALKPCGVRPDMFKNADNDSFGRVOYCPAEIVART
 GCRNDGDNFPAYQIPSSSTISSPYNOPTSTSTSTSTSPVPPTTPSGCTAERMAQ
 CGGNWSGCTTCAAGSTCTKINDYHQL"
 CDS
 1..927
 /codon_start=1
 /product="endoglucanase"
 /protein_id="CA01698.1"
 /db_xref="GI:832898"
 /translation="MRSSPLPSAVVAALPVLALADGRSTRYWDCKPSCGAKKAP
 VNQPVSCNANFORITDFDAKSGCEPGVAVSCADQTPMVAVNDFAFGAFATSIAGSN
 EAGWCACYELEFTSGPVAGKKNVOSTSTGSDLSNHFIDINIGGAGVIGDCTPQF
 GCIPIGRYGGIISRNEDDRPDALKPCGVRPDMFKNADNDSFGRVOYCPAEIVART
 GCRNDGDNFPAYQIPSSSTISSPYNOPTSTSTSTSTSPVPPTTPSGCTAERMAQ
 CGGNWSGCTTCAAGSTCTKINDYHQL"
 BASE COUNT 190 a 377 c 288 g 205 t
 ORIGIN

Query Match 37.3%; Score 437.4; DB 6; Length 1060;
 Best Local Similarity 70.4%; Pred. No. 3.3e-67;
 Matches 655; Conservative 0; Mismatches 246; Indels 30; Gaps 4;

Oy 55 CCAAGATGCGCTACTCCCTCTTCGACACACCTGCGCGCTGCTCCTCTGGTGG 114
 Db 5 CCAAGATGCGCTCTCCCTCCCTCCCTCCCTGCGCGCTGCGCGCTGCGGTGG 64
 Oy 115 CCTCCGCGGCGGCGAGTGGCGACAGTCCAGATGCTGAGTCTGCGACGCGTCT 174
 Db 65 CCCTTGGCGGCTGATGGAG-----GTCCACCGCGCTGCTGAGCTGCTCAAGCCTTCT 118
 Oy 175 GCGCTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 234
 Db 119 GCGGCTGGGCGCAAGAGGCTCCCGTGAACACGCTGTCTTTTCTGCAACGCGCACTTCC 178

Oy 235 AGCGCTGTCCGACTTCAATATGTCAGTGGGCTGCA---CGGCGGCTGGGCTACTCT 291
 Db 179 AGCTATACAGCACTTTCAGACGCCAGTCCGGCTGCGAGCGCGGCGGTGCTGCTACTGCT 238
 Oy 292 GCGCCGACCAAGACTCTCTGGGCGGTGAACAGCAATCTGCTACAGGCTTGGCGGAGCA 351
 Db 239 GCGCCGACCAAGACCCATGCGGCTGTGAACGAGACTTCCGCTGCTGCTTGGCTGCACT 298
 Oy 352 GCATGCGCGGCGGCGGTCCGAACTCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 411
 Db 299 CTATTGCGCGGCGCAATGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 358
 Oy 412 CCGGTCCGCTGCGCGGCAAGCAATGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 471
 Db 359 CCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 418
 Oy 472 GAATTAACAGTTCGATATGCGATGCTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCA 531
 Db 419 GCAGCAACCACTTCTGATTCACATCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCA 478
 Oy 532 GCTGCGAGTTGGGCGGCGCTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 591
 Db 479 CTCGCCAGTTCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 538
 Oy 592 GCGATTCCTTCCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 651
 Db 539 GCGATGCGTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 598
 Oy 652 AGCGGCAACCGCGAGCTTACGTTCCAGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGT 711
 Db 599 AGCGGCAACCACTTCTGATTCACATCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 658
 Oy 712 GCTGCGGCTGCAAGCGCAACAGAGTCCAGTTCGCTGCTGCTGCTGCTGCTGCTGCTGCT 771
 Db 659 GCACCGGATGCGCGCGCGCAAGAGGAGGCAATTCCTGCGGCGGCGGCGGCGGCGGCGG 718
 Oy 772 GCAACGG-----TGGCAGCGGGAGCGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCT 813
 Db 719 GCACACGCTCTCCGGTCAACACAGCTTACAGCAGACACAGCTGCTCCAGCACT 778
 Oy 814 CGGCGCGAGAG---GTCTCCCGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 870
 Db 779 CGAGCGCGGAGTCCAGCTACAGCTTCCAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCT 838
 Oy 871 GCGGTGCTATCGCTTCAGCGGATGACACACCTGTGTCTGTGCGACCACTCCAGAGT 930
 Db 839 GCGGCGGCAATGGCTGAGGGGCTGCACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 898
 Oy 931 TGAACGACTACTGCTGCAAGTGCCTTAAC 961
 Db 899 TTAATGACTGTGATCAATGACTGCTGTAGAC 929

RESULT 11
 A41658
 LOCUS A41658 1060 bp DNA linear PAT 05-MAR-1997
 DEFINITION Sequence 1 from Patent EP0633311.
 ACCESSION A41658
 VERSION A41658.1 GI:2297284
 KEYWORDS
 SOURCE unidentified.
 ORGANISM unidentified.
 REFERENCE 1 (bases 1 to 1060)
 AUTHORS Herbots, I.M., and Jansen, M.P.
 TITLE Hydrophobic amines for cellulase stabilization in liquid detergent
 JOURNAL compositions containing anionic surfactant and cellulase
 Patent: EP 0633311-A 1 11-JAN-1995;
 PROCTER & GAMBLE (US)
 COMMENT Other publication CA 2165771 950105.
 FEATURES
 source
 location/Qualifiers
 1..1060
 /organism="unidentified"

QY 532 GCTCGAGTTCGGCGGCTCCCGGCGCTCAATACGGGGGCGATTTGTCGGCGCAGT 591
 Db 479 CTCGCCAGTTCGGGGGCTCCCGGCGCTCAAGCGGATCTGTCGCCGACAGT 538
 QY 592 GCGATTCCTTCGCCGCGCGCTCAAGCCCGGCTGCACTGGGTTGACTGTTCCAGA 651
 Db 539 GCGATGGGTTCCCGGAGCGCCTCAAGCCCGGCTGCTACTGGCGGTTGACGTTCAGA 598
 QY 652 AGCCCGAACAACCGGAGCTTCACTGCAAGAGTGCAGTGGCCCGGAGATCTGGCC 711
 Db 599 ACCCGCAATCCGAGGTTGAGTTCAGTCCGTCAGTTCAGTCCGAGCCGAGTCTGCTC 658
 QY 712 GCTCCGCTCAAGCGAGCAGCAGCTTCACCTCCGCTTTCAGCCCGCCCAACGGGTG 771
 Db 659 GCACCGGATGCGCGGAGCAGCAGCAGTCCGCTGCGGCTCAAGTCCCTCCAGCA 718
 QY 772 GCAACGG-----TGCGACCGGGAGCGCCAGCTGACGTGCGCTGGGT 813
 Db 719 GCACGACCTCTCCGCTGACACGAGCTTACAGCAGCAGCAGCAGCTTCCACACCT 778
 QY 814 CGGCGCAGAC---GTCTCCCGGCGCGGCGGCTGCTGACGCTTCAGAGTGGGCTCAGT 870
 Db 779 CGAGCCCGCAGTCCAGCTCAAGCTCCAGCGGCTGCTGAGAGTGGGCTCAGT 838
 QY 871 GCGGTGCGATCGGCTTCAAGCGAGTCAACCACTGTCTTGGACCACTGCGCAGAGT 930
 Db 839 GCGGCGGCAATGGCTGAGAGCGGCTGCAACCACTGTGCTGGCGAGCACTTCCAGAGA 898
 QY 931 TGAACGACTACTGCTGAGTGGCTCTAAC 961
 Db 899 TTATGACTGTACCATCAGTCCCTGTAGAC 929

RESULT 13
 ARO59002 1060 bp DNA linear PAT 29-SEP-1999
 LOCUS Sequence 7 from patent US 5837847.
 DEFINITION ARO59002
 ACCESSION ARO59002
 VERSION ARO59002.1 GI:5984579
 KEYWORDS

SOURCE Unknown.
 ORGANISM Unclassified.
 REFERENCE 1 (bases 1 to 1060)
 AUTHORS Royer,J.C., Moyer,D.L., Wendy,Y.T. and Shuster,J.R.
 TITLE Non-toxic, non-toxicogenic, non-pathogenic fusarium expression system
 and promoters and terminators for use therein
 JOURNAL Patent: US 5837847-A 7 17-NOV-1998;
 FEATURES Location/Qualifiers
 source 1..1060

BASE COUNT 190 a 377 c 288 g 205 t
 ORIGIN

Query Match 37.3%; Score 437.4; DB 6; Length 1060;
 Best Local Similarity 70.4%; Pred. No. 3.3e-67;
 Matches 655; Conservative 0; Mismatches 246; Indels 30; Gaps 4;

QY 55 CCAGATGCGCTACATCCCTGCTTCCGACAAACCTGGCGGCTGACATTCCTCGTGG 114
 Db 5 CCAAGATGCGCTTCTCCCTCCCTCCGCTCCGCTGCTGCGCGCTGCTGG 64
 QY 115 CCTCCGCGGCGAGTGGCGAGTGGCCAGATGAGTGTCTGCAAGCCCTGCT 174
 Db 65 CCTTGGCGCTGAGTGGCGAG-----GTCCACCGGCTACTGGAGTGTCTGCAAGCCCTGCT 118
 QY 175 GCGCTTGGCGCGGAGGCGCGCTGAGCCAAACCGGTCTACCGGTGCGATGCCACTTCC 234
 Db 119 GCGGCTGGCGCAAGAGGCTCCCTGGAACCAAGCTCTTCTTCTGCAACGCCACTTCC 178
 QY 235 AGCGCGTGTGCGAGCTTCAATGTCAGTGGGCTGCA---CGGCGGCTGGGCTACTCCT 291
 Db 179 AGCGTATCACGCGACTTCCAGCCAGTCCGGGCTGCGAGCGGCGGCTGCTACTCCT 238

QY 292 GCGCGACCGAGACTCCCTGGGCGGTGAGACAGCAATCTCGCTACAGGCTTGGCCGAGCA 351
 Db 239 GCGCCGACCAAGACCCCATGGGCTGTGAGACAGCACTTGGGCTGGTTTGTGTCACCT 298
 QY 352 GCATTCGCGCGGGGTCCGAATCTCTGCTGTGCTGCGCTCTGACGCGCTCACTTACTT 411
 Db 299 CTATTCGCGGACCAATGAGCGGGGTGTGCTGCGCTCTGACGAGCTCACTTCACT 358
 QY 412 CCGGTCCCGTCCGCGGCAAGCAATGTTGTTGAGTCAAGAGACGCGGGCGACCTGG 471
 Db 359 CCGGTCTGTTGCTGGCAAGATGTTGCTTCAAGTCCAGTCCAGCAGTGGGTATCTTG 418
 QY 472 GAATTAACAGTTTGATATGCGCATGCGCGCGCGCGCTGAGGATCTTCAACGGCTGCA 531
 Db 419 GCAGCAACCACTTGCATCTCAACATCCCGGCGCGCGCTGCGATCTTGCAGGATGCA 478
 QY 532 GCTCGAGTTCGGCGGCTCCCGGCGGCTCAATAGCGCGCATTTGCTGCGGCAACAGT 591
 Db 479 CTCGCCAGTTCGGCGGCTCCCGGCGGCTCAAGCGCTACGCGGCTCTGCTCCGCAAGT 538
 QY 592 GCGATTCCTCCCGGCGGCTCAAGCGCGGCTGCAAGTGGCGGTTGACTGTTCCAGA 651
 Db 539 GCGATGCTTCCCGGAGCGCTCAAGCGCGGCTGCTACTGCGCTTGGACTGTTCAAGA 598
 QY 652 ACSCGCAACCCGAGCTTCAAGTTCAGAGAGTGCAGTGGCCCGCGAGATGTTGCC 711
 Db 599 AGCCGCAATTCGAGGTTGAGTTCGCTGAGTTCGCTGAGTGCAGGCGAGCTGCTGCTC 658
 QY 712 GCTCCGCTGCAAGCGCAGCAGCAGTTCCTCCGCTTCAACCCCGGAGCGGTG 771
 Db 659 GCACCGGATGCGCGGCGGCAAGCAGCAACTTCCCTGCTGCTCAAGTCCCTCCAGCA 718
 QY 772 GCAACGG-----TGCGACCGGGAGCGCCAGCTGCACTGCGCTGGGT 813
 Db 719 GCACGACCTCTCCGCTGACACGAGCTTACAGCAGCAGCAGCAGCTTCCACACCT 778
 QY 814 CGGCGCAGAC---GTCTCCCGGCGCGGCGAGTGGCTGACGCTTCAAGATGGGCTCAGT 870
 Db 779 CGAGCCCGCAGTCCAGCTTCAAGCTCCAGCGGCTGCTGAGAGTGGGCTCAGT 838
 QY 871 GCGGTGCGATCGGCTTCAAGCGAGTCAACCACTGTCTGTCGACCACTGCGCAGAGT 930
 Db 839 GCGGCGCAATGGCTGAGAGGCTGCAACCACTGTGCTGCGAGCACTTCCAGAGA 898
 QY 931 TGAACGACTACTTCCAGTGGCTCTAAC 961
 Db 899 TTATGACTGTACCATCAGTCCCTGTAGAC 929

RESULT 14
 ARO72921 1060 bp DNA linear PAT 28-AUG-2000
 LOCUS Sequence 1 from patent US 5948672.
 DEFINITION ARO72921
 ACCESSION ARO72921
 VERSION ARO72921.1 GI:9999684
 KEYWORDS

SOURCE Unknown.
 ORGANISM Unclassified.
 REFERENCE 1 (bases 1 to 1060)
 AUTHORS Patmusen,G., Mikkelson,J.Moslashedler, Schlein,M.,
 Rasmussen,S.Amant., Hagen,E., Hjort,C.Malland, and Hastup,S.
 TITLE Cellulase preparation comprising an endoglucanase enzyme
 JOURNAL Patent: US 5948672-A 1 07-SEP-1999;
 FEATURES Location/Qualifiers
 source 1..1060

BASE COUNT 190 a 377 c 288 g 205 t
 ORIGIN

Query Match 37.3%; Score 437.4; DB 6; Length 1060;
 Best Local Similarity 70.4%; Pred. No. 3.3e-67;

Matches 655: Conservative 0; Mismatches 246; Indels 30; Gaps 4;	
OY 55	CCAGATGCGCTTACTACCTCCGTTCTGACACACCTGGCGGCTGCACTTCTCTGTCG 114
Db 5	CCAAGATGCGTTCTCTCCCTCTCTCCGCTCGGCTGTGTCGCGCCCTGCGGGTGG 64
OY 115	CCCTCCGGGCGCAGTGGAGTGGCCAGTCCACAGATATCTGGAGTGTGCAAGCCGTGT 174
Db 65	CCCTTCCGCGTGTATGGCAG-----GTCCACCCGCTACTGGAGCTGTGCAAGCCCTTGT 118
OY 175	CGCGTTGGCCCGGGAAGCGCCGCTGACAGCCAGCTTACCGCTGGCATCCAACTTCC 234
Db 119	CGCGTGGGCGAAGAGGCTCCGTTGAACAGCTTCTTCTTCTGCAACGCAACTTCC 178
OY 235	AGCGCTGTGCTGCAATGTCAGTGGGCTGCAAA---CGGGGGCTCGGCTACTGCT 291
Db 179	AGCGTATCAGCGATTTGAGCGCCAGTCCGCTCGAGCCGGGGGTGTGCTCTACTGCT 238
OY 292	GGCGGACACAGACTCCCTGGGGGTGAACGAACTTCGCTTACGGCTTCCGCGGACGA 351
Db 239	GGCGGACACAGACCCTATGGGCTGTGAACGAGACTTCCGCTCGGTTTTCGCAACT 298
OY 352	GCATGCGCGGGGGTCCGAATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 411
Db 299	CTATTGCGCGGAGCAATGAGCGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 358
OY 412	CCGCTCCCGTCCGCGGCAAGCAATGGTGTGTCAGTCAAGCAGCAGTGGCGGCACTGG 471
Db 359	CCGCTCCGTTGTGTGGCAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 418
OY 472	GAATTAACAGTTGCAATGTCAGTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 531
Db 419	GCAGCAACCACTTCTGATCAACATCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 478
OY 532	GTGCGCTGCTGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 591
Db 479	CTCCCGCAGTTTGGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 538
OY 592	GCATTCCTTCCCGCGCGCTCAAGCCGCGGCTGCAAGTGGGCTTGTGACTGTTTCCAGA 651
Db 539	GCATTCGCTTCCCGCGCGCTCAAGCCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 598
OY 652	ACGCGCAACACCGAGCTTCACTGTCAGAGTGTGAGTGTGCGCGCGGCGGCGGCGGCGGCGG 711
Db 599	ACGCGCAACATCGAGCTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 658
OY 712	GTTCGGCTGCAAGCGGAGAGTGTGCAAGTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 771
Db 659	GCACCGGATCCCGCGGAGAGAGCGGCAACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 718
OY 772	GCAACG-----TGCAACCGGAGCGCCAGCTGTCAGTGTGCGCTGGGT 813
Db 719	GCACAGCTCTCGGTCAGCAGCGCTTACAGCAACAGCAACAGCAACAGCAACAGCAACAGCT 778
OY 814	CGGCGCAGAC---GTCTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 870
Db 779	CGAGCCCGCAGTGCACAGCTTCAAGCTTCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 838
OY 871	CGCGTGGCATCGGCTTCAAGCGAGTGTGCAAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 930
Db 839	CGCGCGGCAATGCTGTGAGGCGGCTGCAACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 898
OY 931	TGAAGCACTACTGCT 961
Db 899	TTAATGACTGTACATCACTGCT 929

RESULT 15
 ARI63160
 LOCUS ARI63160 1060 bp DNA 1linear PAT 17-OCT-2001
 DEFINITION Sequence 55 from patent US 6270968.
 ACCESSION ARI63160
 VERSION ARI63160.1 GI:16233662

KEYWORDS	
SOURCE	Unknown.
ORGANISM	Unknown.
REFERENCE	1 (bases 1 to 1060)
AUTHORS	Dalb.o slashed.ge,H., Sandal,T., Kaupinen,M.SakarI. and Diderichsen,Balashdige.
TITLE	Method of providing a hybrid polypeptide exhibiting an activity of interest
JOURNAL	Patent: US 6270968-A 55 07-AUG-2001;
FEATURES	Location/Qualifiers
source	1..1060
BASE COUNT	190 a 377 c 288 g 205 t
ORIGIN	
Query Match	
Best Local Similarity 70.4%; Score 437.4; DB 6; Length 1060;	
Matches 655; Conservative 0; Mismatches 246; Indels 30; Gaps 4;	
OY 55	CCAGATGCGCTTACTACCTCCGTTCTGACACACCTGGCGGCTGCACTTCTCTGTCG 114
Db 5	CCAAGATGCGTTCTCTCCCTCTCTCCGCTCGGCTGTGTCGCGCCCTGCGGGTGG 64
OY 115	CCCTCCGGGCGCAGTGGAGTGGCCAGTCCACAGATATCTGGAGTGTGCAAGCCGTGT 174
Db 65	CCCTTCCGCGTGTATGGCAG-----GTCCACCCGCTACTGGAGCTGTGCAAGCCCTTGT 118
OY 175	CGCGTTGGCCCGGGAAGCGCCGCTGACAGCCAAACCGGCTTACGGCTGCGATCCAACTTCC 234
Db 119	CGCGTGGGCGAAGAGGCTCCGTTGAACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 178
OY 235	AGCGCTGTGCTGCAATGTCAGTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 291
Db 179	AGCTATCAGGACTTGTGAGCGGCAAGTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 238
OY 292	GGCGGACACAGACTCCCTGGGGGTGAACGAACTTCGCTTACGGCTTGGCGGCACTGG 351
Db 239	GGCGGACACAGACCCTATGGGCTGTGAACGAGACTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 358
OY 352	GCATTCCTTCCCGCGCGCTCAAGCCGCGGCTGCAAGTGGGCTTGTGACTGTTTCCAGA 411
Db 299	CTATTGCGCGGAGCAATGAGCGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 358
OY 412	CCGCTCCCGTCCGCGGCAAGCAATGGTGTGTCAGTGTGCGCGCGGCGGCGGCGGCGGCGGCGG 471
Db 359	CCGCTCCGTTGTGTGGCAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 418
OY 472	GAATTAACAGTTGCAATGTCAGTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 531
Db 419	GCAGCAACCACTTCTGATCAACATCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 478
OY 532	GTTCGGCTGCAAGCGGAGAGTGTGCAAGTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 591
Db 479	CTCCCGCAGTTTGGGCGGCT 538
OY 592	GCATTCCTTCCCGCGCGCTCAAGCCGCGGCTGCAAGTGGGCTTGTGACTGTTTCCAGA 651
Db 539	GCATTCGCTTCCCGCGCGCTCAAGCCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 598
OY 652	ACGCGCAACACCGAGCTTCACTGTCAGAGTGTGAGTGTGCGCGCGGCGGCGGCGGCGGCGGCGG 711
Db 599	ACGCGCAACATCGAGCTTCACTGCT 658
OY 712	GTTCGGCTGCAAGCGGAGAGTGTGCAAGTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 771
Db 659	GCACCGGATCCCGCGGAGAGAGCGGCAACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 718
OY 772	GCAACG-----TGCAACCGGAGCGCCAGCTGTCAGTGTGCGCTGGGT 813
Db 719	GCACAGCTCTCGGTCAGCAGCGCTTACAGCAACAGCAACAGCAACAGCAACAGCAACAGCT 778
OY 814	CGGCGCAGAC---GTCTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 870

D _b	779	CGACCCGCCGAGCTGCACGGCTAAGCATTCCCAGCGGCTGCACATGGCTGAAGAGTGGGGCTCAGT	838
O _y	871	GGGTGGCATTCGGCTTTACGGGGATGCAACCACTGTGTCTTGCGACACCCTGGCCAAGT	930
D _b	839	GCGCGGCAATGGGTGGAGGCGCGGTGCACACCTGTGGTGGCGAGCACTTGCACGAAGA	898
O _y	931	TGAACGACTACTACTCGCAGTGGCTCTTAAC	961
D _b	899	TTAATGACTGCTACCATCATCAGTCCCTGTAAAC	929

Search completed: February 27, 2003, 19:35:59
Job time : 3189 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2003, 17:47:15 ; Search time 305 Seconds

(without alignments)
8668.348 Million cell updates/sec

Title: US-10-007-521-11

Perfect score: 1174
Sequence: 1 GAGCAGCCACCCCTCAAGCTG.....AAAAAAAAAAAAAAAAAAAA 1174

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_101002:*

1: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1174	100.0	1174	17 AAT39050	CDNA encoding cell
2	1174	100.0	1174	19 AAV39096	Monocomponent endo
3	453.6	38.6	1261	19 AAV23748	Humicola grisea ce
4	451.6	38.5	922	19 AAV15072	Hybrid DNA compri
5	439	37.4	1060	16 AA260178	H. insolens endogl
6	437.4	37.3	1058	13 AA26405	Humicola insolens
7	437.4	37.3	1060	12 AA014856	Humicola insolens
8	437.4	37.3	1060	13 AA025932	Cellulase containe
9	437.4	37.3	1060	13 AA029934	Endoglucanase gene

10	437.4	37.3	1060	13 AA030067	Sequence encoding
11	437.4	37.3	1060	14 AA041732	Dye transfer inhib
12	437.4	37.3	1060	14 AA049941	Endoglucanase enzy
13	437.4	37.3	1060	17 AAT10182	Alkaline endogluca
14	437.4	37.3	1060	19 AAV16102	Humicola insolens
15	437.4	37.3	1060	19 AAV15065	Humicola insolens
16	437.4	37.3	1060	21 AA053334	Endoglucanase nucl
17	435.8	37.1	1060	13 AA026380	Endoglucanase #1.
18	431	36.7	1060	13 AA030072	43kd endoglucanase
19	429.4	36.6	1060	13 AA031181	H. insolens cellul
20	412.8	35.2	922	19 AAV15072	Hybrid DNA compri
21	393.4	33.3	925	19 AAV15076	Hybrid DNA compri
22	390.6	33.3	807	19 AAV16104	Humicola insolens
23	377.2	32.1	1257	19 AAV19378	Humicola insolens
24	377.2	32.1	1257	19 AAV13840	Humicola insolens
25	377.2	32.1	1257	24 ABL59235	Nucleotide sequenc
26	377.2	32.1	1257	24 AAL3251	Humicola insolens
27	377.2	32.1	1257	24 ABL60700	H. insolens EGV nu
28	368.2	31.4	1154	17 AAT39048	CDNA encoding cell
29	366.4	31.2	928	19 AAV15074	Hybrid DNA compri
30	365.6	31.1	927	19 AAV29596	Humicola insolens
31	353.6	30.1	936	18 AAT6541	20K-cellulase gene
32	337.6	28.8	915	19 AAV15075	Hybrid DNA compri
33	337	28.7	894	17 AAT39061	Chimeric endogluca
34	334.8	28.5	927	17 AAT39062	Fusarium oxysporum
35	315.2	26.8	1473	13 AA014857	Fusarium oxysporum
36	315.2	26.8	1473	13 AA026407	Endoglucanase #2.
37	315.2	26.8	1473	13 AA026382	Endoglucanase gene
38	315.2	26.8	1473	13 AA025933	Endoglucanase gene
39	315.2	26.8	1473	13 AA029935	Endoglucanase enzy
40	315.2	26.8	1473	14 AA049942	F. oxysporum endog
41	315.2	26.8	1473	16 AA260179	Fusarium oxysporum
42	315.2	26.8	1473	19 AAV16103	Chimeric endogluca
43	313.6	26.7	885	17 AAT39075	Dye transfer inhib
44	313.6	26.7	1473	14 AA041733	CDNA encoding cell
45	302	25.7	1132	17 AAT39053	

ALIGNMENTS

RESULT 1	
AAT39050	CDNA encoding cellulolytic enzyme #4 of the invention.
ID AAT39050 standard; CDNA: 1174 BP.	
XX	
AC AAT39050;	
XX	
DT 20-MAY-1997 (first entry)	
XX	
DE CDNA encoding cellulolytic enzyme #4 of the invention.	
XX	
KW Cellulase enzyme; endoglucanase; hydrolysis; cellulose; microorganism;	
KW plant cellulase; catalytic region; textile; backstaining; bio-polishing;	
KW stone-washing; cellulose fabric; colour clarification; defibrillation;	
KW cell wall degradation; paper pulp; debarking; fibre modification;	
KW enzymatic de-inking; drainage improvement; ss.	
XX	
OS Thielavia terrestris.	
XX	
FH	
FT CDS	Location/Qualifiers
FT	59..959
FT	/*tag= a
XX	/product= endoglucanase
PN	
XX	W09629397-A1.
XX	
XX	26-SEP-1996.
PD	
XX	
PF 18-MAR-1996;	96WO-DK00105.
XX	
PR 12-FEB-1996;	96DK-0000137.
PR 17-MAR-1995;	95DK-0000272.
PR 08-AUG-1995;	95DK-0000885.

PR 08-AUG-1995: 95DK-0000886.
PR 08-AUG-1995: 95DK-0000887.
PR 08-AUG-1995: 95DK-0000888.
XX
XX
XX (NOVO) NOVO-NORDISK AS.
PI Andersen LN, Ihara M, Kauppinen MS, Lange L, Lassen SF;
PI Nielsen RI, Schnelein M, Takagi S;
XX
XX WPI, 1996-443173/44.
DR P-PSDB; AAW04928.
PT New endo-glucanase enzyme preparations - contg. conserved catalytic
PT regions, useful for treating fabrics, textiles, plant material or
PT paper pulp
XX
XX Claim 72; Page 160-161; 316pp; English.
XX
XX AAT39047-T39054 represent the coding sequences for the enzymes of the
CC invention. The encoded enzymes possess cellulolytic (particularly
CC endoglucanase) activity. Cellulolytic enzymes are involved in the the
CC hydrolysis of cellulose, and are synthesised by a large number of
CC microorganisms and plants. The enzymes of the invention containing the
CC conserved catalytic regions (such as AAW04913) exhibit improved
CC performance, e.g. 50 times higher performance, compared to multiple
CC domain enzymes. The enzymes can be used for the treatment of fabrics or
CC textiles, preferably for preventing backstaining, for bio-polishing or
CC for stone-washing cellulosic fabric. They can also be used to provide
CC colour clarification for laundry. The enzymes can also be used for the
CC degradation or modification of plant material, such as cell walls. They
CC can also be used in the treatment of paper pulp preferably for
CC debarking, defibration, fibre modification, enzymatic de-linking or
CC drainage improvement.
XX
XX Sequence 1174 BP; 243 A; 395 C; 320 G; 216 T; 0 other:

Query Match 100.0%; Score 1174; DB 17; Length 1174;
Best Local Similarity 100.0%; Pred. No. 1.1e-203;
Matches 1174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

0Y 1 GAGCAGACACCCCTCAAGCTGACAGTTTCACACCCGCTCTCTTTCTTGCGCCCGCAGA 60
1 GAGCAGACACCCCTCAAGCTGACAGTTTCACACCCGCTCTCTTTCTTGCGCCCGCAGA 60
DB 1 GAGCAGACACCCCTCAAGCTGACAGTTTCACACCCGCTCTCTTTCTTGCGCCCGCAGA 60
0Y 61 TGGCCTCTACTCCGCTCTCTTGCGACCAACCGCTGCGCTGCACTTCTCTGCGCTCCG 120
1 TGGCCTCTACTCCGCTCTCTTGCGACCAACCGCTGCGCTGCACTTCTCTGCGCTCCG 120
DB 61 TGGCCTCTACTCCGCTCTCTTGCGACCAACCGCTGCGCTGCACTTCTCTGCGCTCCG 120
0Y 121 CGGCGAGTGGCAGTGGCAGTGCAGAGATCTGGAGCTGTGCAAGCGGTGCGCTT 180
1 CGGCGAGTGGCAGTGGCAGTGCAGAGATCTGGAGCTGTGCAAGCGGTGCGCTT 180
DB 121 CGGCGAGTGGCAGTGGCAGTGCAGAGATCTGGAGCTGTGCAAGCGGTGCGCTT 180
0Y 181 GGCGCGGGAAGGCGCGCTGACGCCAACCCTGCTACGCGTGGATGCCAATTCACAGCGCC 240
1 GGCGCGGGAAGGCGCGCTGACGCCAACCCTGCTACGCGTGGATGCCAATTCACAGCGCC 240
DB 181 GGCGCGGGAAGGCGCGCTGACGCCAACCCTGCTACGCGTGGATGCCAATTCACAGCGCC 240
0Y 241 TGTCCGACTTCAATGTCCAGTGGGCTGCAACGGGCGCTCGGCTCACTCTCTGCGGACCC 300
1 TGTCCGACTTCAATGTCCAGTGGGCTGCAACGGGCGCTCGGCTCACTCTCTGCGGACCC 300
DB 241 TGTCCGACTTCAATGTCCAGTGGGCTGCAACGGGCGCTCGGCTCACTCTCTGCGGACCC 300
0Y 301 AGACTCCCTGGGCGGTGAACAGCAATCTCGCTTACGCGCTTCGCGCGAGAGCATCGCGC 360
1 AGACTCCCTGGGCGGTGAACAGCAATCTCGCTTACGCGCTTCGCGCGAGAGCATCGCGC 360
DB 301 AGACTCCCTGGGCGGTGAACAGCAATCTCGCTTACGCGCTTCGCGCGAGAGCATCGCGC 360
0Y 361 GCGGCTCCGAATCTCTGCTGCTGCGCTGCTACGCGCTCACTTCCGCTCCG 420
1 GCGGCTCCGAATCTCTGCTGCTGCGCTGCTACGCGCTCACTTCCGCTCCG 420
DB 361 GCGGCTCCGAATCTCTGCTGCTGCGCTGCTACGCGCTCACTTCCGCTCCG 420
0Y 421 TCGCGCGCAAGACAGATGGTGGCAAGCAAGAGCACTGGGCGGACCTGGGAAGTAAC 480
1 TCGCGCGCAAGACAGATGGTGGCAAGCAAGAGCACTGGGCGGACCTGGGAAGTAAC 480
DB 421 TCGCGCGCAAGACAGATGGTGGCAAGCAAGAGCACTGGGCGGACCTGGGAAGTAAC 480
0Y 481 AGTTGATATCGCCATGCGCGCGCGCGCTGGGCACTTCAACGAGCTGCACTGCGAGT 540
1 AGTTGATATCGCCATGCGCGCGCGCGCTGGGCACTTCAACGAGCTGCACTGCGAGT 540

DB 481 AGTTGATATCGCCATGCGCGCGCGCGCTGGGCACTTCAACGAGCTGCACTGCGAGT 540
0Y 541 TCGCGCGCTTCCCGCGCTCAATATAGCGCGCATTTGTGCGCGGACAGATGCAATTCCT 600
1 TCGCGCGCTTCCCGCGCTCAATATAGCGCGCATTTGTGCGCGGACAGATGCAATTCCT 600
DB 541 TCGCGCGCTTCCCGCGCTCAATATAGCGCGCATTTGTGCGCGGACAGATGCAATTCCT 600
0Y 601 TCCCGCGCGCTTCAAGCGCGCGCTGCGAGTGGCGGTTTGAAGTGTTCAGAACCGCACA 660
1 TCCCGCGCGCTTCAAGCGCGCGCTGCGAGTGGCGGTTTGAAGTGTTCAGAACCGCACA 660
DB 601 TCCCGCGCGCTTCAAGCGCGCGCTGCGAGTGGCGGTTTGAAGTGTTCAGAACCGCACA 660
0Y 661 ACCCGAGTTTCAAGTTTCAAGAGTGGAGTGGCCCGCGAGATGTTGCGCTCCGCT 720
1 ACCCGAGTTTCAAGTTTCAAGAGTGGAGTGGCCCGCGAGATGTTGCGCTCCGCT 720
DB 661 ACCCGAGTTTCAAGTTTCAAGAGTGGAGTGGCCCGCGAGATGTTGCGCTCCGCT 720
0Y 721 GCAAGCGCAACGAGACTTCCAGCTTCCCGCTTCCAGCTTCCAGCTTCCAGCTTCCAGCT 780
1 GCAAGCGCAACGAGACTTCCAGCTTCCCGCTTCCAGCTTCCAGCTTCCAGCTTCCAGCT 780
DB 721 GCAAGCGCAACGAGACTTCCAGCTTCCCGCTTCCAGCTTCCAGCTTCCAGCTTCCAGCT 780
0Y 781 GCAAGCGCAACGAGACTTCCAGCTTCCCGCTTCCAGCTTCCAGCTTCCAGCTTCCAGCT 840
1 GCAAGCGCAACGAGACTTCCAGCTTCCCGCTTCCAGCTTCCAGCTTCCAGCTTCCAGCT 840
DB 781 GCAAGCGCAACGAGACTTCCAGCTTCCCGCTTCCAGCTTCCAGCTTCCAGCTTCCAGCT 840
0Y 841 GTGCTGCAAGCTTCAAGAGTGGGCTCAGTGGGCTGCGCTTCCAGCTTCCAGCTTCCAGCT 900
1 GTGCTGCAAGCTTCAAGAGTGGGCTCAGTGGGCTGCGCTTCCAGCTTCCAGCTTCCAGCT 900
DB 841 GTGCTGCAAGCTTCAAGAGTGGGCTCAGTGGGCTGCGCTTCCAGCTTCCAGCTTCCAGCT 900
0Y 901 CCGTGTCTTGGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 960
1 CCGTGTCTTGGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 960
DB 901 CCGTGTCTTGGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 960
0Y 961 CAGCTTTTCCAGAGAGTGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1020
1 CAGCTTTTCCAGAGAGTGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1020
DB 961 CAGCTTTTCCAGAGAGTGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1020
0Y 1021 TTTTGGAGCGCTCAAT 1080
1 TTTTGGAGCGCTCAAT 1080
DB 1021 TTTTGGAGCGCTCAAT 1080
0Y 1081 TCTCAGCAGCAGCTTGGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1140
1 TCTCAGCAGCAGCTTGGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1140
DB 1081 TCTCAGCAGCAGCTTGGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1140
0Y 1141 AA 1174
1 AA 1174
DB 1141 AA 1174

RESULT 2
AAV39096
ID AAV39096 standard; cDNA; 1174 BP.
XX
AC AAV39096;
XX
DT 21-SEP-1998 (first entry)
XX
DE Monocomponent endoglucanase encoding cDNA.
XX
KW Monocomponent endoglucanase; cellulolytic enzyme; garment; abrasion;
KW abraded looking jeans; fungus; Thielavia terrestris; stone-washed; ss.
XX
XX Thielavia terrestris.
XX
XX
FH Key Location/Qualifiers
FT CDS /*tag= a
FT /*product= "monocomponent endoglucanase"
XX
XX
XX EP843041-A1.
XX
XX PD 20-MAY-1998.
XX
XX PF 30-MAY-1997; 97EP-0610021.

DR WPI; 1998-217251/19.

PT	Cellulase enzyme variants - having amino acid changes which improve
PT	properties e.g. activity, sensitivity to surfactants, pH optimum or
PT	stability

PS Example 3; Page 77-78; 115pp; English

The present invention describes a cellulase enzyme variant comprising a catalytic core domain exhibiting cellobolytic activity which is derived from a naturally occurring parental cellulase by amino acid residue substitution, insertion and/or deletion, and (with cellulase numbering): (a) at position 5 has an Ala, Ser, or Thr residue; (b) at position 8 has a Phe or a Tyr residue; (c) at position 9 has a Phe, Trp, or Tyr residue; (d) at position 10 has a Asp residue; (e) at position 121 has a Asp residue. The present sequence represents the CDNA sequence of Humicola grisea cellulase from the present invention. The new cellulase enzymes can be used in e.g. detergent or fabric softener compositions, for biopolishing of new fabrics, for promoting a stone-washed look to cellulose containing fabric, for pulp and paper applications, e.g. for debarking, delamination, fibre modification, drainage improvement, inter fibre bonding or for degradation of plant material e.g. for improving feed value. The cellulase variants have improved properties with respect to e.g. catalytic activity, altered sensitivity to anionic tensides, pH optimum or activity profile and stability.

SQ Sequence 1261 BP; 246 A; 417 C; 336 G; 262 T; 0 other;

Query Match	Score	DB	Length
38.68;	453.6;	19;	1261,

Best Local Similarity 74.9%; Pred. No. 2.8e-73;
Matches 583; Conservative 0; Mismatches 189; Indels 6; Gaps 1

Oy	26	TTTTCCACCCGCTCTCTTTTCTTCGCGGCCAGAGTGGCGCTACACCGGCTTCGAC	85
Db	18	TTTTATTACGGCTATTATTATCTTAAACCTCAATATGGCGCTGCTCTATTATTCCGAC	77
Oy	86	AACCTGGCGGCTCACTTCTCTGGTGGCTCCGGCGAGTGGGAGTGGGCACTGCAC	145
Db	78	GGCCCTGGCGGCTCGCGTCCCGCTTGGCGGACACGCGCGGATGGG-----AGTGGAC	131
Oy	146	GAGATACGGGACACGACGCAAGCGGTGGCTGGCGTTGGCCCGGGAAAGCGCGCTACGCA	205
Db	132	CAGATACGGGACACGCTGACAGCCATCTGTCTTTGGCCCGGAAAGCACTCGTGAACA	191
Oy	206	ACCGGTACACCGCTGCGCATGTCACATTCAAGCGCTGTGCACTTCAATGTCCAGTGGG	265
Db	192	GCGTGTCTTCACTTGGCAGCGCAATTTCCAGCGATCAGCGACCCCAATACCAAGTGGG	251
Oy	266	CTGCACAGCGGCTCGGCTTACTCTCTGCGCGACACACATCTCCTGGGCGGTGAACACAA	325
Db	252	CTGCATATGGCGCTCGCGCTTTTGTGTGACCAACCCCTGGGCGCTTGAACACAA	311
Oy	326	TCTGCCATACGGCTTCGCGCGGAGACATGCGCGGCGGTGCGAAATCCTCTGTGTCTG	385
Db	312	TCTGCCATATGGCTTCCGTGCGCACGCGTATTTGGGTTGATTCGGAACCTCGTGTGTCTG	371
Oy	386	CGCTGTACAGCGCTACACTTCCGCTCCGTCGCGCGGCAAGACAATGTGTGCA	445
Db	372	CGCATGTACGCTGTACTTTACACTCGGCGGCTGTGGCGGCAAGACATGTGTCTGCA	431
Oy	446	GTCACAGACACTGCGCGCGACCTGGGAATACCAATTTGGATATCGCCATGCGCGGCGG	505
Db	432	GTTCACCAACACCGCGCGGATCTCGGACAGCAACATTTTGGACTCCAGATTTCCAGGCGG	491
Oy	506	CGGCGTGGGCACTTCAACGGCTGAGCTGCAGTTGGGGGCGCTCCCGGCGCTCAATA	565
Db	492	CGGTGCGGCACTTTGTATGGGTGCAACCCCAATTTGGAGATCTCCGTGGCAACGCTA	551
Oy	566	CGCGGCAATTTGTGCGCGGACACAGTGCATTTCCGCCGCGCGCTCAAGCCGGCTG	625
Db	552	CGGTGCAATCTAGACCGAGTCTCTGGACTCGTTCCGTCGCGCGCTCAAGCCGGCTG	611
Oy	626	CCAGTGGCGGTTTGACTGGTTCCAGACCGCGACACCCGACGTTTCCAGTTCACGACAGT	685

[illegible]

RESULT 4
2A015073

ID	AAV15072 standard; DNA; 922 BP
----	--------------------------------

AC AAV15072;

DT 19-MAY-1998 (first entry)

Hybrid DNA comprising a family 45 cellulase core region.

KW Cellulase core region; isolation; microorganism; identification;

XX
2
1
1
2
The

OS Chimeric - Humicola insolens.

Key

FT
FT

XX
XX
DN

XX
DD 30-NOV-1997

XX 12-MAY-1997. 97WQ-DK00216.
DE

10-MAY-1996: 96DK-0000562

XX
PA (NOVO) NOVO-NORDISK AS.

XX Diderichsen B, Kauppinen S, Sandal T;
PT

XX WPT: 1998-008878/01.
DB

DK P-PSDB; AA44203
XX

PI Isolating novel DNA sequences from microorganisms
PT for culturing the microorganism

Example 3; Page 51-52; 72pp; English.

The present sequence represents a novel hybrid gene construct from an example of the present invention. The present invention describes a novel method for providing a novel DNA sequence encoding a polypeptide from a microorganism with an activity of interest. The method comprises: (i) PCR amplification of the DNA with PCR primers with homology to (a) known gene(s) encoding a polypeptide with an activity of interest; (ii) linking the obtained PCR product of a 5' structural gene sequence and a 3' structural gene sequence; (iii) expressing the resulting hybrid DNA sequence; (iv) screening for hybrid DNA sequences encoding a polypeptide with the activity of interest or a related activity; and (v) isolating the hybrid DNA sequence identified in step (iv). This method provides for identification and isolation of sequences from microorganisms without having to cultivate and isolate the microorganism.

sq Sequence 922 BP; 151 A; 333 C; 262 G; 176 T; 0 other;

Query Match	38.5%	Score 451.6;	DB 19;	Length 922;
Best Local Similarity	71.6%;	Pred: No. 6,2e-73;		
Matches 630; Conservative	0;	Mismatches 229;	Indels 21;	Gaps 2

PN EP495554-A.
 XX 22-JUL-1992.
 PD 15-JAN-1992; 92EP-0200101.
 PF 16-JAN-1991; 91EP-0870006.
 PR 06-NOV-1991; 91EP-0202879.
 PR 06-NOV-1991; 91EP-0202881.
 XX (PROC) PROCTER & GAMBLE CO.
 PA Baack AC, Boulisque J, Bush A, Convents AC, Hargrove RS;
 PI Pretlyaj, Busch A, Ceulemans RAA;
 XX WPI: 1992-243405/30.
 DR P-PSDB; AAR25428.
 XX
 PT Detergent compsn. for improving cleaning and performance - is
 PT composed of quat. ammonium cpd. and high activity cellulose e.g.
 PT homogeneous endo:glucanase
 PS Claim 10; Page 19; 28pp; English.
 XX
 CC The cellulase DNA was used in a detergent compsn. contg. a
 CC quaternary ammonium cpd. The compsn. gives a detergent with good
 CC cleaning and softening performance due to the synergistic effect of
 CC the ammonium and cellulase components. The cellulase is a
 CC homogeneous endoglucanase component which is immunoreactive with an
 CC antibody raised against a cellulase purified from Humicola isolens
 CC DSM 1800. See also AAQ25933.
 CC
 SQ Sequence 1060 BP; 190 A; 377 C; 288 G; 205 T; 0 other;

Query Match 37.3%; Score 437.4; DR 13; Length 1060;
 Best Local Similarity 70.4%; Pred. No. 2,4e-70;
 Matches 655; Conservative 0; Mismatches 246; Indels 30; Gaps 4;

QY 55 CGAGATGGCTCTACTCCGCTTCTCGACAAACCTGGCGGCTGCACTTCTGTGTCG 114
 DB 5 CCAAGATGGCTTCTCTCCCTCTCTCCGCTGCGGCTGCGGCTGCGGCTGCGG 64
 QY 115 CCTCCGCGCAGTGGCAGTGGCAGTGCAGATGATGAGTGGTGTGCAAGCGTGT 174
 DB 65 CCTCCGCGCAGTGGCAGTGGCAGTGCAGATGATGAGTGGTGTGCAAGCGTGT 118
 QY 175 GCGCTTGGCCCGGGAAGCGCGCTGACGCAACGCGTCTACGCGTGGATGCCACTTCC 234
 DB 119 GCGGCTGGGCGCAAGAGGCTCCGCTGAACCGAGCTGTCTTCTCTGCAAGCGCAACTTCC 178
 QY 235 AGGCGCTGTCCGCTCAATGTCAGTGGGCTGCA---CGGCGGCTCGGCTACTGCT 291
 DB 179 AGGCTATCACGGACTTTCGAGCCCAAGTCCGCTGCGAGCGGCGGCTGTGCTTCTGCT 238
 QY 292 GCGCGCAGCAGACTTCCCTGGCGGTGAACGACAATCTCGGCTTTCGCGCGAGCA 351
 DB 239 GCGCGCAGCAGACTTCCCTGGCGGTGAACGACAATCTCGGCTTTCGCGCGAGCA 298
 QY 352 GCATCGCGCGGCGGTGCGATCTCTGCTGCTGCGGCTGCTGCTGCTGCTGCTGCTGCT 411
 DB 299 CTATTCGCGCGCAATGTCAGTGGGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCT 358
 QY 412 CGGCTCCGCTGCGCGCAAGACAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 471
 DB 359 CGGCTCCGCTGCTGCGCAAGACAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 418
 QY 472 GAAGTACCAAGTTCGATATCGCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 531
 DB 419 GCAGCAACACACTTCGATCTCAACATCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 478
 QY 532 GCTCGCAGTTGCG 591
 DB 479 CTCCCGAGTTGCG 538

QY 592 GCGATTCCTTCCCGCGCGCGCTCAAGCCCGGCTGCCAGTGGCGGTTTGACTGTTCCAGA 651
 DB 539 GCGATCGGTTCCCGCGCGCGCGCTCAAGCCCGGCTGCCAGTGGCGGTTTGACTGTTCCAGA 598
 QY 652 ACCCGCAACCCGAGCTTTCAGAGTTCAGAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGT 711
 DB 599 ACCCGCAACCCGAGCTTTCAGAGTTCAGAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGT 658
 QY 712 GCTCGCGCTGCGCAAGCGCAAGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 771
 DB 659 GCGCGGATGCGCGCGCGCGCAAGCGCAAGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 718
 QY 772 GCAACGG-----TGGCAGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 813
 DB 719 GCACGAGCTTCCCGGTCAGACAGCTACAGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 778
 QY 814 CGGCGCAGAC---GTCCTCCG 870
 DB 779 CGAGCGCGCGCAGTCCAGCTTCCAGTCCAGCTTCCAGCTTCCAGCTTCCAGCTTCCAGCT 838
 QY 871 GCGGTGGCATGCGCTTTCAGCGGATGCAACCACTGTGCTTCTGCGCAGCAGCAGTGCAGAGT 930
 DB 839 GCGCGGCAATGCTGAGAGCGGCTGACACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 898
 QY 931 TGAACGACTACTTTCGCGAGTGCCTCTTAAC 961
 DB 899 TTATGACTGCTACCATGATGCTGCTAGAC 929

RESULT 9
 AAQ29934
 ID AAQ29934 standard; cDNA; 1060 BP.
 XX
 AC AAQ29934;
 DT 09-MAR-1993 (first entry)
 XX
 DE Endoglucanase gene.
 XX
 KW Alkaline cellulase; laundry detergent compositions;
 KW fabric colour maintenance; ss.
 XX
 OS Humicola insolens DSM 1800.
 XX
 FH Key Location/Qualifiers
 FT CDS 10..927
 FT sig_peptide 10..72
 FT mat_peptide 73..924
 FT /*tag= c
 XX
 PN EP508358-A.
 XX
 PD 14-OCT-1992.
 XX
 PF 07-APR-1992; 92EP-0105956.
 PR 12-APR-1991; 91EP-0870062.
 PR (PROC) PROCTER & GAMBLE CO.
 PA Busch A, MacCorquodale F;
 PI WPI: 1992-341667/42.
 DR P-PSDB; AAR27968.
 XX
 PT Laundry detergent compsn. - contg. alkali cellulase and PVP for
 PT fabric colour maintenance upon laundering
 PS Disclosure; Page 14; 23pp; English.
 XX

FH	Key	Location/Qualifiers
227		10 007

175 GCGCTTGGCCCGGGAAGGCCCGCGCTAAGCCAAACCGTCTACGCGTGGATGCCAACTTCC 23

FH	key	location/qualifiers
FT		10..927
CDS		

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2003, 18:29:56 : Search time 1642 Seconds

(without alignments)
11579.477 Million cell updates/sec

Title: US-10-007-521-11

Perfect score: 1174
Sequence: 1 GAGCAGACACCCCTCAAGCTG.....AAAAAAAAAAAAAAAAAAAA 1174

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapept 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estrov:*
6: em_estrpl:*
7: em_estro:*
8: em_hic:*
9: gb_estl:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estcom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_huv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	213	18.1	691	10	BE585661 EST#6SP6
2	130.4	11.1	450	13	BI200729 clfosf.r
3	114.4	9.7	444	13	BI190695 13910fs.r
4	106.4	9.1	426	13	BI187295 alh11fs.r
5	70.2	6.0	168	14	BO152857 NF025H04I
6	69.6	5.9	203	9	AI620089 ty48d08.x

Result No.	Score	Query Match	Length	ID	Description
7	68.8	5.9	493	12	BG56882 df25g08.y
8	68.4	5.8	441	13	BM573394 f34f05.y
9	68.4	5.8	596	14	BQ186959 ut-e-EJ1
10	68	5.8	179	9	AA237797
11	68	5.8	369	9	AI870238 mx14e11.r
12	68	5.8	442	9	AI721722
13	67.6	5.8	271	10	AM193007 x168h01.x
14	67.6	5.8	424	14	BO391956
15	67.6	5.8	441	12	BG897070
16	67.6	5.8	451	14	BO525205
17	67.6	5.8	476	13	BI496243
18	67.6	5.8	542	14	BO522269
19	67.6	5.8	543	14	BO521817
20	67.6	5.8	546	14	BO520637
21	67.6	5.8	550	14	BO396262
22	67.4	5.7	273	9	AU175573
23	67.4	5.7	435	9	AA806378
24	67.2	5.7	191	14	BO942645
25	67.2	5.7	201	9	AL712405
26	67.2	5.7	215	14	BO932982
27	67.2	5.7	246	9	AL598823
28	67.2	5.7	400	10	AV760389
29	67.2	5.7	400	10	AV760391
30	67.2	5.7	640	12	BF342223
31	67	5.7	248	9	AA738097
32	67	5.7	295	9	AA268010
33	67	5.7	335	9	AA270150
34	67	5.7	353	10	AA335461
35	67	5.7	376	12	BG653511
36	67	5.7	379	9	AA023416
37	67	5.7	399	10	AW088521
38	67	5.7	435	14	BO043166
39	67	5.7	479	13	BM114912
40	67	5.7	501	13	BM116326
41	67	5.7	698	12	BF136434
42	67	5.7	740	12	BG294200
43	67	5.7	786	12	BG297249
44	67	5.7	1279	12	BE784196
45	66.8	5.7	215	13	BI190568

ALIGNMENTS

RESULT 1
BE585661
LOCUS
DEFINITION
EST#6SP6_D02_d2_014 KSU wheat Fusarium graminearum infected spike
CDNA library Trilicium aestivum CDNA clone EST#6SP6_D02_d2_014,
mRNA sequence.
ACCESSION
BE585661
VERSION
BE585661.1 GI:9838604
KEYWORDS
EST.
SOURCE
bread wheat.
ORGANISM
Trilicium aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
; Triliceae; Trilicium.
1 (bases 1 to 691)
Fellers,J.P., Li,W.L., Hill-Ambroz,K., Matthews,A. and Gill,B.S.
The structure and function of the expressed portion of the wheat
genomes - Kansas State University. Fusarium graminearum infected
spike CDNA library
Unpublished (2000)
Contact: John Fellers
US Department of Agriculture, Agriculture Research Service, Plant
Science and Entomology Unit
Dept. of Plant Pathology, 4006 Throckmorton Hall, Kansas State
University, Manhattan, KS 66506, USA
Tel: 785-532-2367
Fax: 785-532-6167
Email: jpf@alfalfa.ksu.edu
Sequence have been trimmed to remove vector sequence and low

JOURNAL COMMENT

Sequence have been trimmed to remove vector sequence and low

LOCUS	168 bp	mrna	linear	EST 24-Apr-2000
DEFINITION	B0152857 NF025H04IR1P1044 Irradiated Medicago truncatula cDNA clone			
ACCESSION	B0152857			
VERSION	B0152857.1			
KEYWORDS	GI:20289916			
SOURCE	EST.			
ORGANISM	barrel medic. Medicago truncatula			
REFERENCE	Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons, core eudicots; Rosidae, eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.			
AUTHORS	1 (bases 1 to 168) Torres-Jerez,I., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J., Flores,H.R., Inman,J.T., Weller,J.W. and May,G.D.			
TITLE	Expressed Sequence Tags from the Samuel Roberts Noble Foundation Medicago truncatula irradiated library			
JOURNAL	Unpublished (2001)			
COMMENT	Contact: May GD Plant Biology Division The Samuel Roberts Noble Foundation 2510 Sam Noble Parkway, Ardmore, OK 73402, USA Tel: 580 221 7391 Fax: 580 221 7380 Email: gmay@noble.org Insert Length: 168 Std Error: 0.00 Plate: 025 row: H column: 04 Seq primer: TCACACAGGAACGCTATGCAC. Location/Qualifiers 1..168 /organism="Medicago truncatula" /db_xref="taxon:3880" /clone="NF025H04IR" /clone_1id="Irradiateds" /tissue_type="seedlings" /dev_stage="seedling" /note="Vector: lambda Zap; Seedlings were exposed either to 100 Gy gamma or 0.5, 1, 5, or 10 kJ/m2 UV irradiation. Gamma-irradiated samples were harvested at 6, 12, 24 and 48 hours after treatment. UV-irradiated samples were harvested 24 hours post-treatment. cDNA was prepared from polyA+ enriched, pooled samples of equivalent amounts of total RNA from each sample. The cDNA was directionally ligated into the Uni-Zap XR vector (Stratagene) and packaged using the Gigapack III Gold packaging extracts. Phagemids containing cDNA inserts were in vivo excised from the recombinant Uni-Zap XR vector using Exsist helper phage and the E. coli strain XL1-Blue MR' (Stratagene). Excised plasmids were plated using SOLR cells."			
BASE COUNT	85 a	16 c	24 g	43 t
ORIGIN				
Query Match	6.0%; Score 70.2; DB 14; Length 168;			
Best Local Similarity	75.7%; Pident.No. 2.3;			
Matches	87; Conservative 0; Mismatches 28; Indels 0; Gaps 0;			
OY	1060	TACATGACGCGCGGTACATCTCACACGACCTTTGGGGCGGAATCAGCGCGCTTTTA	1119	
Db	54	TACATCATTAATGTTGGAAATATATATGATCGATGGATTTGGTGATGTTTAACTCTTTTAA	113	
OY	1120	AA	1174	
Db	114	AA	168	
RESULT 6				
LOCUS	A1620089			
DEFINITION	tY48d08.x1 NCI_CGAP_Ut2 Homo sapiens cDNA clone IMAGE:2282319 3', mrna sequence.			
ACCESSION	A1620089			
VERSION	A1620089.1			
GI	4629215			

KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
TITLE	1 (bases 1 to 203)
JOURNAL	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .
COMMENT	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index Unpublished (1997)
FEATURES	Contact: Robert Strausberg, Ph.D. Email: cgapbs-rt@mail.nih.gov Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNLW at: www-bio.11n1.gov/bdrrp/image/image.html Insert Length: 2710 Std Error: 0.00 Seq primer: -40UP from Gibco High quality sequence stop: 163 POLYA-No.
Source	Location/Qualifiers 1..203 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:2282319" /clone_lib="NCI CGAP ut2" /tissue_type="moderately-differentiated endometrial adenocarcinoma, 3 pooled tumors" /lab_host="DH10B" /note="Organ: uterus; Vector: pCMV-Sport6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.85 kb. Life Technologies catalog #: 11539-012"
BASE COUNT	35 a 48 c 33 g 87 t
ORIGIN	
Query Match	5.9%; Score 69.6; DB 9; Length 203;
Best Local Similarity	81.0%; Pred. No. 2.5;
Matches	81; Conservative 0; Mismatches 19; Indels 0; Gaps 0;
OY	1075 TACACATCTGCACCGCATTTGGGGGGCGAATCCGCCGTTTAAAAA 1134
Db	126 TAAAAACCCCGGGCCCTTGGGGGCCCTTTCCCTTTTAAAAA 67
OY	1135 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1174
Db	66 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 27
RESULT 7	
LOCUS	BG656882 493 bp mRNA linear EST 26-APR-2001
DEFINITION	df25908.Y1 wellcome CRC pBN3 St10 5 Xenopus laevis cDNA clone
ACCESSION	IMAGE:3558351 5', mRNA sequence.
VERSION	BG656882
KEYWORDS	BG656882.1 GI:13796412
SOURCE	EST.
ORGANISM	African clawed frog. Xenopus laevis
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae; Xenopodinae; Xenopus.
AUTHORS	1 (bases 1 to 493) Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D., Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person , B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R., Waterston,R. and Wilson,R.
TITLE	Washu Xenopus EST Project, 1999
JOURNAL	Unpublished (1999)

MEDLINE
COMMENT

97044477
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
CDNA Library preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).
The following repetitive elements were found in this CDNA sequence: 299-359 >POLY_A#Simple_repeat
Seq primer: M13 REVERSE.

FEATURES
SOURCE

Location/Qualifiers
1. 596
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="UI-E-EJ1-ajy-j-05-0-UI"
/clone_lib="UI-E-EJ1"
/tissue_type="fetal eyes, lens, eye anterior segment, optic nerve, retina, Retina Foveal and Macular, RPE and Choroid"
/dev_stage="fetal and adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site_1: EcoR I; Site_2: Not I; UI-E-EJ1 is a subcloned CDNA library constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand CDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded CDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand CDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tags for this library are: fetal eyes, AGAATCAAGA; lens, CGATTAGCGA; eye anterior segment, AATGCCGCAAT; optic nerve, CCATTAGTG; retina, CCGCG; Retina Foveal and Macular, GTCC; RPE and Choroid, ACCCTA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."
BASE COUNT
191 a 144 c 125 g 136 t
ORIGIN

Query Match 5.8%; Score 68.4; DB 14; Length 596;
Best Local Similarity 69.4%; Pred. No. 2;
Matches 93; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 1041 ACATACCTTGATCTTGATACATAGCAGCCGGTACACATCTCAGCCGACTTTGGGG 1100
11 111 111 111 11 11 111 11 11 11
DB 220 ATATCTGCTTCCTCTCTCTCTCTCCAGCTCAGACATCAGACATCAGACTTTTCAT 279
11 111 111 111 111 111 111 111 111 111
QY 1101 CGGAATCGAGCCGCTTTAAAAA 1160
11 111 111 111 111 111 111 111 111 111
DB 280 TACAGTCAGCTATGTTTAAAAA 339
11 111 111 111 111 111 111 111 111 111
QY 1161 AAAAAAAAAAAAAA 1174
11 111 111 111 111 111 111 111 111 111
DB 340 AAAAAAAAAAAAAA 353

RESULT 10
AA237797 179 bp mRNA linear EST 03-MAR-1997
LOCUS mxi4e11.r1 Soares mouse NMU Mus musculus cDNA clone IMAGE:680204
DEFINITION 5', mRNA sequence.
ACCESSION AA237797
VERSION AA237797.1 GI:1861820
KEYWORDS EST.

SOURCE
ORGANISM

house mouse.
Mus musculus
Mammalia: Eutheria: Rodentia: Scurionath: Muridae: Murinae: Mus.
REFERENCE
1 (bases 1 to 179)
Maira, M., Hillier, J., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Scheinberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.

TITLE
JOURNAL
COMMENT
The WashU-HMI Mouse EST Project
Unpublished (1996)
Contact: Maira M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: mouseest@wustl.edu
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:419908
Seq primer: -26m13 rev2 ET from Amersham
High quality sequence stop: 166.

FEATURES
SOURCE

Location/Qualifiers
1. 179
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:680204"
/clone_lib="Soares mouse NMU"
/tissue_type="liver"
/lab_host="DH10B"
/note="Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand CDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAAGTGGAGGCGCCGCAATCTTTTTTT 3']; double-stranded CDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library constructed and normalized by Bento Soares and M.Fatima Bonaldo."
BASE COUNT
85 a 29 c 21 g 44 t
ORIGIN

Query Match 5.8%; Score 68; DB 9; Length 179;
Best Local Similarity 80.0%; Pred. No. 4.5;
Matches 80; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 1075 TACACATCTCACACGACTTTGGGGCGGAATCAGGCCGTTTAAAAA 1134
11 111 111 111 11 11 11 11 11 11 11 11 11 11 11 11
DB 72 TCCACGGCTTTATGACATTAAGGTGATGATGACACTTGA 131
11 111 111 111 111 111 111 111 111 111
QY 1135 AAAAAAAAAAAAAA 1174
11 111 111 111 111 111 111 111 111 111
DB 132 AAAAAAAAAAAAAA 171

RESULT 11
A1870238/c 369 bp mRNA linear EST 07-MAR-2000
LOCUS w120e09.x1 NCI-CGAP_Ut1 Homo sapiens cDNA clone IMAGE:2425480 3',
DEFINITION mRNA sequence.
ACCESSION A1870238
VERSION A1870238.1 GI:5544206
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
Mammalia: Eutheria: Primates: Catarrhini: Hominiidae: Homo.

REFERENCE 1 (bases 1 to 369)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index

REFERENCE AUTHORS TITLE	JOURNAL COMMENT
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo, 1 (bases 1 to 271) NCI-CCAF http://www.ncbi.nlm.nih.gov/ncicgap . National Cancer Institute, Cancer Genome Anatomy Project (CGAP) Tumor Gene Index	Unpublished (1997)
Contact: Robert Strausberg, Ph.D.	

BASE COUNT	51 a	71 c	56 g	93 t
ORIGIN				

Query Match	5.8%;	Score 67.6;	DB 10;	Length 271;
Best Local Similarity	80.6%;	Pred. No. 4;		
Matches 79;	Conservative 0;	Mismatches 19;	Indels 0;	Gaps 0;
Qy 1077	CACATCTCACACGCACTTGGGGGCGGGAATCAGCCGCTTTAAAAA	AAAAAAAAAAAAAAAAAAAA	1136	
Db 130	CCCCCCCCCCCCCTTTGGGGGCTGGTGGCCCTTGGGGA	AAAAAAAAAAAAAAAAAAAA	71	
Qy 1137	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	AAAAAAAAAAAA	1174	
Db 70	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	AAAAA	33	

RESULT 14	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
BO391956/c	BO391956	424 bp mRNA linear EST 22-MAY-2002	NIHSC:mq21f07.x2	NICHD XGC Emb5	Silurana tropicalis cDNA clone	western clawed frog.	Silurana tropicalis
			INSC:5309316 3,	mRNA sequence.			
			BO391956				
			BO391956.1	GI:21079643			
			EST.				

FEATURES

```
source 1. .424
/compare="c:\nurs +rod\c\ic
```

BASE COUNT	117 a	76 c	68 g	163 t	
ORIGIN					
Query Match		5.8%;	Score 67.6;	DB 14;	Length 424;
Best Local Similarity	64.9%;		Pred. No. 3.1;		
Matches 100;	Conservative 0;	Mismatches 54;	Indels 0;	Gaps 0;	
QY 1019	ATTTTTGGAGCGCTCAATACATACATTAACCTTTCGATTTCTTGACTATAGCAGCCCGGTACA				1078
Db 155	AGTTTCTTAAATTCGCGGGGTAATAAAAAAACTTAATCCTTGACACCTAATAATGTAAATTTT				96
QY 1079	CATCTCACACCGACACTTGGGGGCGGAATCAGGCGCCGTTTAAAAAATAAAAAATAAAAA				1138
Db 95	TAACTATCATATAAACTGGGGAGGAGGAGGAAAAAATAAAAAATAAAAAATAAAAA				36
QY 1139	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA				1172
Db 35	AAAAAAAAAAAAAAAAAAAAAAAAAAAAA				2

RESULT	15
LOCUS	BG897070
DEFINITION	BG897070 441 bp mRNA linear EST 06-NOV-2001 H0445-1-03.R HOA (Human Osteoarthritic Cartilage) Homo sapiens cDNA
ACCESSION	BG897070
VERSION	BG897070
KEYWORDS	, mRNA sequence.
SOURCE	BG897070.1 GI:14307311
ORGANISM	EST. human. Homo sapiens

LOCUS	B6897070	441 bp	mRNA	linear	EST 06-NOV-2001
DEFINITION	HOA45-1-03. R HOA (human Osteoarthritic Cartilage) Homo sapiens CDNA				
ACCESSION	B6897070				
VERSION	B6897070.1				
KEYWORDS	GI:14307311				
SOURCE	EST.				
ORGANISM	human.				
REFERENCE	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 441)				
AUTHORS	Kumar,S., Connor,J.R., Dadds,R.A., Halsey,W., Van Horn,M., Mao,J., Seth,G., Mui,P., Agarwal,P., Badger,A.M., Lee,J.C., Gowen,M. and Lark,M.W.				
TITLE	Identification and initial characterization of 5000 expressed sequenced tags (ESTs) each from adult human normal and osteoarthritic cartilage cDNA libraries				
JOURNAL	Osteoarthr. Cartil. 9 (7), 641-653 (2001)				
MEDLINE	21482651				
COMMENT	Contact: Sanjay Kumar				

FEATURES	Location/Qualifiers
source	1. .441

BASE COUNT	177 a	94 c	106 g	64 t
ORIGIN				

FEATURES

Query Match 5.8%; Score 67.6; DB 12; Length 441;
 Best Local Similarity 77.4%; Pred. No. 3;
 Matches 82; Conservative 0; Mismatches 24; Indels 0; Gaps 0;
 QY 1069 CGCCGTTACATCTCACACCGACTTTGGGGCGGGAATCAGCCCGTTTAAAAAAA 1128
 Db 268 CGCCACTGCACTCCGAGTCGGCGTGGCGCAGAGCGAGACTCCGTCCTCAAAAAAAA 327
 QY 1129 AA 1174
 Db 328 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 373

Search completed: February 27, 2003, 20:03:37
 Job time : 1650 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2003, 17:41:10 : Search time 137 Seconds

(without alignments)
5344.003 Million cell updates/sec

Title: US-10-007-521-11

Perfect score: 1174

Sequence: 1 GAGCAGCACCCCTCAAGCTG.....AAAAAAAAAAAAAAAAAAAA 1174

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 460893 seqs, 311809382 residues

Total number of hits satisfying chosen parameters: 921786

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_NA.*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	437.4	37.3	1060	US-09-735-787-1	Sequence 1, Appl
2	353.6	30.1	936	US-08-841-636A-30	Sequence 30, Appl
3	315.2	26.8	1473	US-09-735-787-3	Sequence 3, Appl
4	83.2	7.1	887	US-08-841-636A-36	Sequence 36, Appl
5	66	5.6	682	US-09-764-847-20	Sequence 20, Appl
6	65.6	5.6	371	US-09-925-239-16	Sequence 16, Appl
7	64.6	5.5	1341	US-09-764-846-41	Sequence 41, Appl
8	64	5.5	95	US-09-919-580-246	Sequence 246, Appl
9	63.8	5.4	1908	US-09-745-763-69	Sequence 69, Appl
10	63.6	5.4	917	US-09-925-301-29	Sequence 29, Appl
11	63.4	5.4	1377	US-09-822-830A-588	Sequence 588, App
12	63.2	5.4	355	US-09-960-352-14757	Sequence 14757, A
13	63	5.4	1545	US-09-764-877-3172	Sequence 3172, A
14	62.6	5.3	391	US-09-960-352-14316	Sequence 14316, A
15	62.6	5.3	1817	US-09-764-864-197	Sequence 1, Appl
16	62.6	5.3	2203	US-10-224-562-1	Sequence 1, Appl
17	62.6	5.3	2203	US-09-801-861-1	Sequence 1, Appl
18	62.2	5.3	2379	US-10-036-041-71	Sequence 71, Appl
19	62.2	5.3	2379	US-10-035-835-71	Sequence 71, Appl

20	62.2	5.3	2379	9	US-10-174-590-483	Sequence 483, App
21	62.2	5.3	2379	9	US-10-176-758-483	Sequence 483, App
22	62.2	5.3	2379	9	US-10-175-737-483	Sequence 483, App
23	62.2	5.3	2379	9	US-10-173-706-483	Sequence 483, App
24	62.2	5.3	2379	9	US-10-175-738-483	Sequence 483, App
25	62.2	5.3	2379	9	US-10-175-752-483	Sequence 483, App
26	62.2	5.3	2379	9	US-10-176-482-483	Sequence 483, App
27	62.2	5.3	2379	9	US-10-176-757-483	Sequence 483, App
28	62.2	5.3	2379	9	US-10-176-913-483	Sequence 483, App
29	62.2	5.3	2379	9	US-10-180-552-483	Sequence 483, App
30	62.2	5.3	2379	9	US-10-180-557-483	Sequence 483, App
31	62.2	5.3	2379	9	US-09-931-836-71	Sequence 71, Appl
32	62.2	5.3	2379	9	US-10-173-700-483	Sequence 483, App
33	62.2	5.3	2379	9	US-10-174-572-483	Sequence 483, App
34	62.2	5.3	2379	9	US-10-174-579-483	Sequence 483, App
35	62.2	5.3	2379	9	US-10-174-582-483	Sequence 483, App
36	62.2	5.3	2379	9	US-10-174-588-483	Sequence 483, App
37	62.2	5.3	2379	9	US-10-175-739-483	Sequence 483, App
38	62.2	5.3	2379	9	US-10-175-740-483	Sequence 483, App
39	62.2	5.3	2379	9	US-10-176-488-483	Sequence 483, App
40	62.2	5.3	2379	9	US-10-176-492-483	Sequence 483, App
41	62.2	5.3	2379	9	US-10-176-747-483	Sequence 483, App
42	62.2	5.3	2379	9	US-10-176-750-483	Sequence 483, App
43	62.2	5.3	2379	9	US-10-176-985-483	Sequence 483, App
44	62.2	5.3	2379	9	US-10-176-987-483	Sequence 483, App
45	62.2	5.3	2379	9	US-10-176-987-483	Sequence 483, App

ALIGNMENTS

RESULT 1
US-09-735-787-1
Sequence 1, Application US/09735787
Patent No. US20010036910A1
GENERAL INFORMATION:
APPLICANT: Rasmussen, Grethe
Mikkelsen, Jan Moller
Schulein, Martin
Patkar, Shankant A.
Hagen, Fred
TITLE OF INVENTION: A Cellulase Preparation Comprising an
Endoglucanase Enzyme
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESSEE: NO. US20010036910A1 No. US20010036910A1disk of No. US2001001
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/735,787
FILING DATE: 13-Dec-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/189,028
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Lambitis, Elias J.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 3469,214-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO. 1:
SEQUENCE CHARACTERISTICS:
LENGTH: Humicola insolens

LENGTH: 936 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Melanocarpus alomyces
STRAIN: ALKO4237
FEATURE:
NAME/KEY: exon
LOCATION: 33..115
OTHER INFORMATION: /codon_start= 33
OTHER INFORMATION: /product= "20k-cellulase"
FEATURE:
NAME/KEY: exon
LOCATION: 187..435
OTHER INFORMATION: /product= "20k-cellulase"
FEATURE:
NAME/KEY: exon
LOCATION: 506..881
OTHER INFORMATION: /product= "20k-cellulase"
US-08-841-636A-30

Query Match 30.1%; Score 353.6; DB 7; Length 936;
Best Local Similarity 72.8%; Pred. No. 5.4e-57;
Matches 505; Conservative 0; Mismatches 119; Indels 70; Gaps 1;

145 CGAGATACCTGGGACTGCTGGAAGCCCTGCGCTTGCGCCGGGAAGCCGCCCTCAGCC 204
183 CCGAGTACTGGAGTCTGCTGAGCCGCTGCGCTGCGCCGGGAAGCCGCCCTGAGAAC 242
205 AACCGCTACGCGTGGGATGCGCAACTCCAGCGCTGCGCACTTCAATGTCAGTCGG 264
243 AGCCGCTACTGCTGGAGGCGCAACTCCAGCGCTGCGCACTTCAATGTCAGTCGG 302
265 GCTGCAAGCGCGCTGCGCTTCTCTGCGCCGACAGACTCCCTGGCGGTGAAGACA 324
303 GCTGCGAGGCGCGCGCGCTTCTCTGCGCCGACAGACTCCCTGGCGGTGAAGACA 362
325 ATCTGCGCTACGCGTGGGATGCGCAACTCCAGCGCTGCGCACTTCAATGTCAGTC 384
363 ACCTTCTGTAAGCGCTTCTGCGCGCACTGCACTGAGCGCCAGAGAGTCTGTGCT 422
385 GCGCTGCTACCGG-----CTGACCTTCACTTCCGCTCCGCTCGCGGCAAGACA 434
423 GTGCTGCTACGCGTGGGATGCGCAACTCCAGCGCTGCGCACTTCAATGTCAGTCAG 482
399 -----CTGACCTTCACTTCCGCTCCGCTCGCGGCAAGACA 434
483 TGACCGAGCGACCGCTCGCGCACTTGGGAGTAAACAGCTTCAATGTCAGTCAGTC 494
435 ATGGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 494
543 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 602
495 ATGCGCGCGCGCGCGCTGCGCACTTCAAGCGCTGCGCACTTCAAGCGCTGCGCGC 554
603 ATCCCGCGCGCGCGCGCTGCGCACTTCAAGCGCTGCGCACTTCAAGCGCTGCGCGC 662
555 GCGCGCTACATGCGCGCGCGCTGCGCACTTCAAGCGCTGCGCACTTCAAGCGCTGCG 614
663 GCGCGCTACATGCGCGCGCGCTGCGCACTTCAAGCGCTGCGCACTTCAAGCGCTGCG 722
615 AAGCGCGCGCGCGCGCTGCGCACTTCAAGCGCTGCGCACTTCAAGCGCTGCGCGC 674
723 AAGCGCGCGCGCGCGCTGCGCACTTCAAGCGCTGCGCACTTCAAGCGCTGCGCGC 742
675 TTTCAGCAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 734
783 TTTCAGCAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 842
735 GACTCAGGCTGCGCGCTGCGCACTTCAAGCGCTGCGCACTTCAAGCGCTGCGCGC 768

Db 843 GACGCGGCTTCCGCGCTTCAAGCGCCCGCAGCG 876

RESULT 3

US-09-735-787-3
Sequence 3, Application US/09735787
Patent No. US20010036910A1

GENERAL INFORMATION

APPLICANT: Rasmussen, Grethe
Mikkelsen, Jan Moller
Schulein, Martin
Parker, Shankant A.
Hagen, Fred

TITLE OF INVENTION: A Cellulase Preparation Comprising an
Endoglucanase Enzyme

NUMBER OF SEQUENCES: 33

CORRESPONDENCE ADDRESS:
ADDRESSEE: No. US20010036910A10 No. US20010036910A1dlsk of No. US200100

STREET: 405 Lexington Avenue, 64th Floor

CITY: New York

STATE: New York

COUNTRY: United States of America

ZIP: 10174-6401

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/735,787

FILING DATE: 13-Dec-2000

CLASSIFICATION: <Unknown>

Prior Application Data:
APPLICATION NUMBER: 09/189,028

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J.

REGISTRATION NUMBER: 33,728

REFERENCE/DOCKET NUMBER: 3469, 214-US

TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123

TELEFAX: 212-878-9655

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 1473 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

MOLECULE TYPE: cdna

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:
ORGANISM: Fusarium oxysporum

STRAIN: DSM 2672

FEATURE:
NAME/KEY: CDS

LOCATION: 97..1224

SEQUENCE DESCRIPTION: SEQ ID NO: 3:

US-09-735-787-3

Query Match 26.8%; Score 315.2; DB 10; Length 1473;
Best Local Similarity 68.5%; Pred. No. 7.3e-50;

Matches 466; Conservative 0; Mismatches 208; Indels 6; Gaps 2;

83 CACAACCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 142
108 CACTCTTCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 167
143 CAGGATATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAG 202
168 TACTCATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAG 227
203 CCAACGGTCTACGCGTGCATGCCAATTCAGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 262

```

Db 228 CGCCCTGCTTTACTTGTATAGAACGACAAACCCATTTCACACCAATGCTGTCAA 287
Oy 263 GGGCTGCAAA--CGGGGCTGCGGCTACTCTCGCGGACAGACTCCCTGGGGGTGAA 319
Db 288 CGATTGTGAGGGGTGATGTTCTGCTATGCTTGACCACTACTCTCCCTGGGCTGCAA 347
Oy 320 CGACATCTGCGCTAGGGCTGCGCGGAGAGACATCGCGGGGGCCGAGATCCTGTG 379
Db 348 CGATTGAGCTTGTGCTAGCGTTGCTGCTACCAAGATCTCGGTGGCTCGAGGCGAGCTG 407
Oy 380 GATGCTGCGCTGTACGCGCTCACTTCCGCTCGGCTCGCGGCAAGACATGAT 439
Db 408 GTCCTGTGCTGTATGCTTGTGACCTTACACACTGCGCCGCTCAAGGCAAGAGATGAT 467
Oy 440 GGTGACGTCAAGACAGACTGGGCGGACCTGGGAGATGACCACTTGCATGCTGCTCC 499
Db 468 CGTCCAGTCCACCAACACTGAGAGTGTATCGGCGACACCACTTGCATGATGATCC 527
Oy 500 CGGCGGCGGCTGGGATCTTCAACGGCTGAGCTGCGAGTGGGCG--GGCTCCCGCG 556
Db 528 CGGCGGTGGTGGTGTCTGACGCGCTGACCTGTGAGTGGCAAGGCTCTCGGCGG 587
Oy 557 CGCTCAATACGGCGGCTATTCGTGCGCGACCACTGCGATTCCTTCCCGCGCGCTCAA 616
Db 588 TGCCAGTACGGGGGTATCTCTCCGAGAGATGTGATGATACCCGAGCTTCTCAA 647
Oy 617 GCCCGGCTGCACTGGCGGTTTACTGTTTCAAGAGCCGCAACCCGAGCTTCACTT 676
Db 648 GGCAGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 707
Oy 677 CGACAGGTGCTGAGTGGCGGCGGAGATGTTGCGGCTCGGCTGCAAGGCGCAAGCA 736
Db 708 TGAGCAGGTTCAGTGGCCCGGAGAGCTGCTCTGACATGATGAGCAAGGCTGATGAGA 767
Oy 737 CTCAGCTTCCCGCTTCTCA 756
Db 768 CTCAGCTTCCCGCTTCTCA 787

```

```

RESULT 4
US-08-841-636A-36
; Sequence 36, Application US/08841636A
; Patent No. US20020168751A1
; GENERAL INFORMATION:
; APPLICANT: Miettinen-Oinonen, Arja
; APPLICANT: Londesborough, John
; APPLICANT: Vehmaanper, Jari
; APPLICANT: Haakana, Hei
; APPLICANT: M ntyl, Arja
; APPLICANT: Lantto, Raija
; APPLICANT: Elovaio, Minna
; APPLICANT: Joutsio, Vesa
; APPLICANT: Paloheimo, Marja
; TITLE OF INVENTION: NOVEL CELLULASES, THE GENES ENCODING THEM AND
; TITLE OF INVENTION: USES THEREOF
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, N.W., Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/841,636A
; FILING DATE: 30-APR-1997

```

```

; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/005,335
; FILING DATE: 17-OCT-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/007,926
; FILING DATE: 04-DEC-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/020,840
; FILING DATE: 28-JUN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/732,181
; FILING DATE: 16-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FI96/00550
; FILING DATE: 17-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Timothy J. Shea, Jr.
; REGISTRATION NUMBER: 41,306
; REFERENCE/DOCKET NUMBER: 1716.0510005/MAC/TJS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)371-2600
; TELEFAX: (202)371-2540
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 887 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Melanocarpus albus
; STRAIN: ALKO4237
; FEATURE:
; NAME/KEY: exon
; LOCATION: 351..455
; OTHER INFORMATION: /product= "protein-with-CBD"
US-08-841-636A-36

```

```

Query Match 7.1%; Score 83.2; DB 7; Length 887;
Best Local Similarity 73.6%; Pred. No. 4.9e-07;
Matches 106; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

```

```

Oy 817 GCCAGAGCTCTCCGCGGCGGCGAGTGTGCTGACGCTTCAGAACTGGGCTCAGTGGCGTG 876
Db 313 GCAACCGGCGCAACGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 372
Oy 877 GATGCGGCTTCAAGGATGACCACTGTGTCTGTGACACACCTGCGCAGAGTTGAAG 936
Db 373 GATGCGGCTTCAAGGATGACCACTGTGTCTGTGACACACCTGCGCAGAGTTGAAG 932
Oy 937 ACTACTACTGCGAGTGGCTTAA 960
Db 433 ACTACTACTGCGAGTGGCTTAA 456

```

```

RESULT 5
US-09-764-847-20
; Sequence 20, Application US/09764847
; Patent No. US20020132767A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009
; CURRENT APPLICATION NUMBER: US/09/764,847
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION DATA: 2001-01-17
; PRIOR APPLICATION DATA: removed - consult PALM or file wrapper
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 20
; LENGTH: 682
; TYPE: DNA
; ORGANISM: Homo sapiens

```

US-09-764-847-20

Query Match 5.6%; Score 66; DB 10; Length 682;
Best Local Similarity 83.3%; Pred. No. 0.00071;
Matches 75; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

OY 1085 ACACCGACTTGGGGCGGAATCAGGCCGTTTAAAAA
1144
DB 563 ACACCGACTTGGGGCGGAATCAGGCCGTTTAAAAA
622
OY 1145 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1174
DB 623 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 652

RESULT 6

US-09-925-299-16
; Sequence 16, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
; LENGTH: 371
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (350)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (360)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-299-16

Query Match 5.6%; Score 65.6; DB 10; Length 371;
Best Local Similarity 60.8%; Pred. No. 0.00075;
Matches 107; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

OY 999 CCGTCACCTTCGTCATATTTTGGAGCGCTCAATACATACATTAACCTTCGATTCTT 1058
DB 169 CCAACCCCTTATATTTGTAATGATGGCTGAACATTTGTCCTGTTCTACTT 228
OY 1059 GTACATAGCAGCGGTCACATCTCAGACGACTTGGGGCGGAGTCAAGCCGTTT 1118
DB 229 CTACCACTAGCTTTTACCAATTAAGCTCTACTGTCTGAAAAA 288
OY 1119 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1174
DB 289 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 344

RESULT 7

US-09-764-846-41
; Sequence 41, Application US/09764846
; Patent No. US20020102638A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT212
; CURRENT APPLICATION NUMBER: US/09/764,846
; PRIOR FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 348

SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 41
; LENGTH: 1341
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-846-41

Query Match 5.5%; Score 64.6; DB 10; Length 1341;
Best Local Similarity 83.9%; Pred. No. 0.0014;
Matches 73; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

OY 1088 CCGACTTGGGGCGGAATCAGGCCGTTTAAAAA
1147
DB 1217 CAGACCTGGCAACAGACTGAGACCCCTGTCACAAAAA
1276
OY 1148 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1174
DB 1277 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1303

RESULT 8

US-09-919-580-246/c
; Sequence 246, Application US/09919580
; Patent No. US20020110832A1
; GENERAL INFORMATION:
; APPLICANT: Pyle, Ruth
; APPLICANT: Xu, Jiangchun
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.552
; CURRENT APPLICATION NUMBER: US/09/919,580
; CURRENT FILING DATE: 2001-07-30
; NUMBER OF SEQ ID NOS: 934
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 246
; LENGTH: 95
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 73, 78
; OTHER INFORMATION: n = A,T,C or G
US-09-919-580-246

Query Match 5.5%; Score 64; DB 10; Length 95;
Best Local Similarity 97.0%; Pred. No. 0.0012;
Matches 64; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1109 GGCCGTTTAAAAA
1168
DB 83 GGCCGTTTAAAAA
24
OY 1169 AAAAAA 1174
DB 23 AAAAAA 18

RESULT 9

US-09-745-763-69
; Sequence 69, Application US/09745763
; Patent No. US20020065394A1
; GENERAL INFORMATION:
; APPLICANT: McCoy, Kenneth
; APPLICANT: Jacobs, Kenneth
; APPLICANT: Lavalie, Edward R.
; APPLICANT: Collins-Racie, Lisa A.
; APPLICANT: Evans, Cheryl
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Spaulding, Vilki
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; ENCODING THEM

```

QY      1089 CGACCTTGGGGCGGGAATCAGCCCGTTTAAAAAAAAAAAAAAAAAAAAAAAAA 1148
Db      814 CTACTCTGTGGCAATTAACACACAATTAGCAAAAAAAAAAAAAAAAAAAAAA 873
QY      1149 AAAAAAAAAAAAAAAAAAAAAAAAAA 1174
Db      874 AAAAAAAAAAAAAAAAAAAAAA 899

RESULT 11
US-09-822-830A-588
: Sequence 588, Application US/09822830A
: Patent No. US20020142952A1
: GENERAL INFORMATION:
: APPLICANT: Genetics Institute, Inc.
: APPLICANT: Wong, Gordon G.
: APPLICANT: Clark, Hilary
: APPLICANT: Fechtel, Kim
: APPLICANT: Agostino, Michael J.
: APPLICANT: Howes, Steven H.
: APPLICANT: Resnick, Richard J.
: APPLICANT: Gulinkota, Kamalakara
: APPLICANT: Graham, James R.
: TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
: FILE REFERENCE: GIN 6402
: CURRENT APPLICATION NUMBER: US/09/822,830A
: CURRENT FILING DATE: 2001-03-29
: PRIORITY APPLICATION NUMBER: 60/195,504
: PRIORITY FILING DATE: 2000-04-06
: PRIORITY NO.: 631

```

```

? SOFTWARE: PatentIn Ver. 2.0
? SEQ ID NO 588
? LENGTH: 1377
? TYPE: DNA
? ORGANISM: Homo sapiens
US-09-822-830A-588

Query Match      5.4%; Score 63.4; DB 10; Length 1377;
Best Local Similarity 82.0%; Pred. No. 0.0024;
Matches 73; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 1086 CACCGACTTTGGGGCGCGGAATCAGGCCCGTTTAAAAAAAAAAAAAAAAAAAA 1145
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1262 CCGTCGATTTGGGGAGAAATAAACAGATATGAGTTTAAAAAAAAAAAAAAAA 1321
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      1146 1147 1148 1149 1150 1151 1152 1153 1154 1155 1156 1157 1158 1159 1160

```

```

D6      1322  AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA  1330

RESULT  12
US-09-960-352-14757/c
? Sequence 14757, Application US/09960352
? Patent No. US20020137139A1
?
? GENERAL INFORMATION:
? APPLICANT: Warren, Wesley C.
? APPLICANT: Tao, Nengding
? APPLICANT: Byatt, John C.
? APPLICANT: Mathilaan, Nagappan
? TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
? FILE REFERENCE: 16511.006/37-21(10298)C
? CURRENT APPLICATION NUMBER: US/09/960,352
? CURRENT FILING DATE: 2001-09-24
?
? NUMBER OF SEQ ID NOS: 15112
?
? SEQ ID NO 14757
?
? LENGTH: 355
? TYPE: DNA
? ORGANISM: Bos taurus
? OTHER INFORMATION: Clone ID: 63-LIB3058-014-Q1-K1-H4

```


US-09-960-352-14757

Query Match

Best Local Similarity 5.4%; Score 63.2; DB 10; Length 355;
Matches 65; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1107 CAGCCCGCTTTAAAAA

DB 147 CGGCCCGCTTTAAAAA

QY 1167 AAAAAAA 1174

DB 87 AAAAAAA 80

RESULT 13

US-09-764-877-3172/C
Sequence 3172, Application US/09764877
Patent No. US20020147140A1

GENERAL INFORMATION:

APPLICANT: Rosen et al.

TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

FILE REFERENCE: PC005

CURRENT APPLICATION NUMBER: US/09/764,877

CURRENT FILING DATE: 2001-01-17

Prior application data removed - refer to PALM or file wrapper

NUMBER OF SEQ ID NOS: 4031

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 3172

LENGTH: 1545

TYPE: DNA

ORGANISM: Homo sapiens

US-09-764-877-3172

Query Match

Best Local Similarity 5.4%; Score 63; DB 10; Length 1545;
Matches 75; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 1079 CATCTCACACCGACTTGGGGCGGGAATCAGCCGCTTTAAAAA

DB 843 CATCTCACACCGACTTGGGGCGGGAATCAGCAATTCGCTCAAAAAA

QY 1139 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1173

DB 783 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 749

RESULT 14

US-09-960-352-14316/C
Sequence 14316, Application US/09960352
Patent No. US20020137139A1

GENERAL INFORMATION:

APPLICANT: Warren, Wesley C.

APPLICANT: Tao, Nengding

APPLICANT: Byatt, John C.

APPLICANT: Mathialagan, Nagappan

TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND

FILE REFERENCE: 16511.006/37-21(10298)C

CURRENT APPLICATION NUMBER: US/09/960,352

CURRENT FILING DATE: 2001-09-24

NUMBER OF SEQ ID NOS: 15112

SEQ ID NO 14316

LENGTH: 391

TYPE: DNA

ORGANISM: Bos taurus

OTHER INFORMATION: Clone ID: 61-LIB3058-052-Q1-K1-H10

US-09-960-352-14316

Query Match

Best Local Similarity 5.3%; Score 62.6; DB 10; Length 391;
Matches 68; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1098 GGGCGGAATCAGCCCGCTTTAAAAA

DB 137 GGGCGCAATGTGGCGCGA

QY 1158 AAAAAAAAAAAAAA 1174

DB 77 AAAAAAAAAAAAAA 61

RESULT 15

US-09-764-864-197
Sequence 197, Application US/09764864
Patent No. US20020132753A1

GENERAL INFORMATION:

APPLICANT: Rosen et al.

TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

FILE REFERENCE: PT23

CURRENT APPLICATION NUMBER: US/09/764,864

CURRENT FILING DATE: 2001-01-17

Prior application data removed - consult PALM or file wrapper

NUMBER OF SEQ ID NOS: 1792

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 197

LENGTH: 1817

TYPE: DNA

ORGANISM: Homo sapiens

US-09-764-864-197

Query Match

Best Local Similarity 5.3%; Score 62.6; DB 10; Length 1817;
Matches 74; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 1082 CTCACACCGACTTGGGGCGGGAATCAGCCGCTTTAAAAA

DB 1718 CTCCTCCTGACAAATGAATGTTCTGCAAAAAA

QY 1142 AAAAAAAAAAAAAAAAAAAAAA 1174

DB 1778 AAAAAAAAAAAAAAAAAAAAAA 1810

Search completed: February 27, 2003, 18:37:26
Job time: 147 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 24, 2003, 11:42:49 ; Search time 85 Seconds
(without alignments)
468.729 Million cell updates/sec

Title: US-10-007-521-12

Perfect score: 1659
Sequence: 1 MRSTPYLRTTLAALPLVAS.....TTCVSGTTCQKLDYYSQCL 299

Scoring table: BL0SUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_101002:.*
1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1659	100.0	299	17	AAW04928
2	1659	100.0	299	19	AAW63624
3	1178	71.0	305	19	AAW44854
4	1178	71.0	305	12	AAW41929
5	1159	69.9	305	12	AAW45271
6	1159	69.9	305	13	AAW25525
7	1159	69.9	305	13	AAW25464
8	1159	69.9	305	13	AAW25428
9	1159	69.9	305	13	AAW27968
10	1159	69.9	305	13	AAW28295

11	1159	69.9	305	14	AAW37150
12	1159	69.9	305	14	AAW42063
13	1159	69.9	305	16	AAW67388
14	1159	69.9	305	16	AAW01502
15	1159	69.9	305	17	AAW88471
16	1159	69.9	305	19	AAW46616
17	1159	69.9	305	21	AAW44266
18	1159	69.9	305	21	AAW03660
19	1146	69.1	305	17	AAW83352
20	1145	69.0	305	17	AAW83356
21	1143	68.9	305	17	AAW83359
22	1142.5	68.9	306	19	AAW44269
23	1142	68.8	305	13	AAW28300
24	1142	68.8	305	17	AAW83355
25	1142	68.8	305	17	AAW83358
26	1138	68.6	305	17	AAW83353
27	1137	68.5	305	17	AAW83354
28	1137	68.5	305	17	AAW83357
29	1137	68.5	305	17	AAW83360
30	1134	68.4	305	13	AAW28818
31	1127	67.9	286	19	AAW57420
32	1122	67.6	200	19	AAW53968
33	1120	67.5	284	15	ABW04137
34	1119	67.5	284	15	ABW04129
35	1117	67.3	200	19	AAW53979
36	1114	67.1	284	15	ABW04141
37	1113	67.1	284	15	ABW04140
38	1112	67.0	284	15	ABW04128
39	1112	67.0	284	15	ABW04130
40	1110	66.9	305	17	AAW83361
41	1108	66.8	284	15	ABW04135
42	1108	66.8	284	15	ABW04135
43	1107.5	66.8	306	19	AAW44270
44	1107	66.7	284	15	ABW04132
45	1106.5	66.7	267	15	ABW04136

ALIGNMENTS

RESULT 1
AAW04928
ID AAW04928 standard; Protein: 299 AA.
XX
AC AAW04928;
XX
DT 20-MAY-1997 (first entry)
XX
DE Cellulytic enzyme #4 of the invention.
XX
KW Cellulytic enzyme; endoglucanase; hydrolysis; cellulose; microorganism;
KW plant cellulase; catalytic region; textile; backstaining; bio-polishing;
KW stone-washing; cellulosic fabric; colour clarification; defibrillation;
KW cell wall degradation; paper pulp; debarking; fibre modification;
KW enzymatic de-inking; drainage improvement.
XX
OS Thielavia terrestris.
XX
FN WO9629397-A1.
XX
PD 26-SEP-1996.
XX
PF 18-MAR-1996; 96WO-DK00105.
XX
PR 12-FEB-1996; 96DK-0000137.
PR 17-MAR-1995; 95DK-0000272.
PR 08-AUG-1995; 95DK-0000885.
PR 08-AUG-1995; 95DK-0000886.
PR 08-AUG-1995; 95DK-0000887.
PR 08-AUG-1995; 95DK-0000888.
XX
PA (NOVO) NOVO-NORDISK AS.
XX

Dye transfer inhib
Endoglucanase enzy
H. insolens endogl
43 kd endoglucanase
Alkaline endoglucan
Humicola insolens
Endoglucanase amin
Humicola insolens
Humicola insolens
H. insolens mutant
Hybrid DNA protein
43kd endoglucanase
Humicola insolens
Humicola insolens
Humicola insolens
Humicola insolens
H. insolens cellul
Thielavia terrestr
Mutant 43kd endogl
Mutant 43kd endogl
Thielavia terrestr
Humicola insolens
Mutant 43kd endogl
Humicola insolens
Mutant 43kd endogl
Hybrid DNA protein
Mutant 43kd endogl

[illegible]

Dd	239	SSPVGQPTSTSTSTSSPPVQTTPSGCTAERMAOGGGMGSGCTTCVAGSTCTKIN	298
OY	293	DYISQCL 299 :	
Dd	299	DMYHOCL 305	
<hr/>			
RESULT 4			
AAW41929	ID	AAW41929 standard; Protein: 305 AA.	
XX	AC	AAW41929;	
XX	AC		
XX	XX	26-JUN-1998 (first entry)	
DT	XX		
DE	XX	Humicola insolens endoglucanase cellulase NCE4.	
XX	XX		
KW	XX	Endoglucanase: cellulase; NCE4: Fluff removal; weight reduction:	
KX	XX	denter reduction; bleaching; denim dyed fibre.	
XX	XX		
OS	XX	Humicola insolens.	
FH	XX		
FT	XX	Key location/Qualifiers	
FT	XX	Peptide 1..22	
FT	XX	/label= sig_peptide	
FT	XX	Peptide 23..305	
FT	XX	/label= mat_peptide	
PN	XX		
XX	XX	WO9803640-A1.	
PD	XX		
XX	XX	29-JAN-1998.	
PF	XX		
PE	XX	24-JUL-1997; 97WO-JP02561.	
PR	XX		
XX	XX	24-JUL-1996; 96JP-0194974.	
PA	XX	(MEIJ) MEIJI SEIKA KAISHA LTD.	
PI	XX	Aoyagi K, Hamaya T, Koga J, Kono T, Moriya T, Murakami T;	
PI	XX	Murashima K, Sumida N;	
DR	XX		
XX	XX	wpl: 1998-120765/11.	
XX	XX	N-P5DB; AAV13840.	
PT	XX		
PT	XX	Endoglucanase cellulase NCE4 from Humicola insolens - for treatment	
XX	XX	of cellulose fibres for fluff elimination and weight reduction	
PS	XX	Claim 1; Pages 33-36; 43pp: Japanese.	
CC	XX		
CC	XX	The present sequence is the Humicola insolens	
CC	XX	endoglucanase cellulase NCE4, which can be used to treat cellulose	
CC	XX	fibres for fluff removal, weight (denier) reduction and bleaching,	
CC	XX	especially of denim dyed fibres.	
SQ	Sequence	305 AA:	
<hr/>			
Query Match		71.0%; Score 1178; DB 19; Length 305;	
Best Local Similarity		68.1%; Pred. No. 1.8e-82;	
Matches 209; Conservative		44; Mismatches 44; Indels 10; Gaps 4;	
OY	1	MRSFVLRTTTLAALPLVASAASGSQGSTRYWDCKPSCAMPKAAVSQPVYACDANFOR 60	
Dd	1	MRSFLRLSRAVAAYLAPLVALLAA--DGKSPRYWDCKPSCGMAKKPVNQVFSCNANFOR 58	
OY	61	LSDFNVOGSCN-GGSAYISCADOTPPAAVNDNLAYGFATSIAGSESMSWCACATLTFTSG 119	
Dd	59	LTFDFAKSGCEGEGGAAYSCADQTPAAVNDFAFGFAATSIAGSENGWCACACELETFTSG 118	
OY	120	PAAGTMTVVVOSTSTGDJGNSOFDIAMPGGGIFNCGSSPGGLPGAQYGISSDDCD 179	
Dd	119	PVAGKKMYYVOSTSTGDDJGNSHFDLNIRGGVGITFDGCIPQTEGGLPGQARTGISRNED 178	
OY	180	SFPABLPCGCCORRDFWFONADNPFTTFQOVCPAEIVARSGKRDDSSFP-VFTPPSGC 238	

	Matches	206; Conservative	44; Mismatches	47; Indels	10; Gaps	4;
QY	1	MRSTVLRRTLLAALPLVLAASAAGSGOSTRRYDDCKPSCAMPKAAVSPVYACANFOR	60			
Db	1	MSSSEPLRLSAVVAALPLVLAAL--DGRSTRYDDCKPSCGMAKKKAPVNGPVESCANFOR	58			
QY	61	LSDFVVGSGCN--GSSAVSCADOTPMNAVNDNLKYGFAATSIAGGSESSWCCACATLFTTSG	119			
Db	59	ITDFPAKSGSGCEPGVAVSCADOTPMNAVNDNFLGFAATSIAGSNAGWCACAEELFTTSG	118			
QY	120	PVAGKTRMYVOSTSTGTGDLGSGNOFDDAMPGCGGVIENGSGSPOFGGLPGAGVGISSDDOC	179			
Db	119	PVAGKTRMYVOSTSTGTGDLGSHNFDLNTIRGGVGITFDGCTPQFGGLPGAGVGISSRNECD	178			
QY	180	SFPAPLKPCCOMRFMDQFONADNPTFTFOOVOCPAEIVARSCKRNLDDSSFP--VFTRPSSG	238			
Db	179	RFPDLAKPGCYWRPFMFKNADNPSFSPFOYQCPALVATGCRNRDDGDFRPAVOQLPSSST	238			
QY	239	NGGTGTPRTSTAPGSGQTS-----PGGSGCTSQKWAQCGTIGSGGCTTVCASGTTQCKLN	292			
Db	239	SSPVNQPPRTSTRTSTTSSPPVQPTPTPSGCTAEERMAWCGGNGMGSGCTTCVAVGSTCTKLN	298			
QY	293	DYYSQCL 299				
Db	299	DHYHQCL 305				

RESULT	7
AAR25464	
ID	AAR25464 standard; Protein: 305 AA.
XX	
AC	AAR25464;
XX	
DT	07-JAN-1993 (first entry)
XX	
DE	Endoglucanase #1.
XX	
KW	CMC-endoase; 43 kD cellulase; monoclonal antibody.
XX	
OS	Humicola insolens.
XX	
FH	Key
FT	peptide
FT	Location/Qualifiers
FT	1..21
FT	/label= signal_peptide
FT	22..305
FT	Protein
FT	/label= Mature_protein
XX	
PN	EP495257-A.
XX	
PD	22-JUL-1992.
XX	
PF	06-NOV-1991; 91EP-0202879.
XX	
PR	16-JAN-1991; 91EP-0870006.
PR	06-NOV-1991; 91EP-0202880.
PR	06-NOV-1991; 91EP-0202879.
XX	
PA	(PROC) PROCTER & GAMBLE CO.
XX	
PI	Baeck AC, Busch A, Ceulemans RAA;
XX	
WI	WPI; 1992-243163/30.
DR	N-PSDB; AAQ26380.
XX	
PT	Compact, granular detergent compens. - contain high activity
PT	cellulase and softening clay to provide synergistic effect in
PT	softening performance
XX	
PS	Disclosure; Page 20-21; 29pp; English.
XX	
CC	The sequences given in AAR25464 and AAR25466 are endoglucanases which
CC	are immunoreactive with a monoclonal antibody raised against a
CC	partially purified 43 kD cellulase derived from Humicola insolens.

CC These endoglucanases exhibit a CMC-endooase activity of at least
CC about 50, pref. at least about 60, imparicular at least about 90 CMC-
CC endooase units per mg of total protein. These endoglucanases have
CC molecular weight of approx. 43 kD.
XX
SQ Sequence 305 AA;

Query Match	69.9%	Score 1159:	DB 13,	Length 305;
Best Local Similarity	67.1%	Pred. No. 5.2e-81;		
Matches 206;	Conservative 44;	Mismatches 47;	Indels 10;	Gaps 4;

[illegible]

XX	RESULT 8
XX	AAAR25428
XX	AAAR25428 standard; Protein; 305 AA.
XX	AAAR25428;
XX	15-JAN-1993 (first entry)
XX	cellulase contained in a detergent compsn.
XX	Endoglucanase; immunoreactive; Humicola isolens; cleaning;
XX	softening.
XX	Humicola isolens.
XX	EP495554-A.
XX	22-JUL-1992.
XX	15-JAN-1992; 92EP-0200101.
XX	16-JAN-1991; 91EP-0870006.
XX	06-NOV-1991; 91EP-0202879.
XX	06-NOV-1991; 91EP-0202881.
XX	(PROC) PROCTER & GAMBLE CO.
XX	Baeck AC, Boutique J, Bush A, Convents AC, Hargrove RS;
XX	Prettyaj, Busch A, Ceulemans RA;
XX	WPI; 1992-243405/30.
XX	P-PSDB; AAAR25428.
XX	Detergent compsn. for improving cleaning and performance - is
XX	composed of quat. ammonium cpd. and high activity cellulose e.g
XX	homogeneous endo:glucanase

XX (NOVO) NOVO-NORDISK AS.
 XX
 XX Tsuchiya R, Wagner P;
 PI
 XX WPI; 1992-382092/46.
 DR
 DR N-PSDB; AAQ30067.
 XX
 PT Detergent additive contg. cellulase and specific protease - which
 PT does not degrade the cellulase during storage and clarifies the
 PT colour of dyed cellulosic materials
 XX
 PS Disclosure; Page 8-9; 15pp; English.
 XX
 CC The ~43 kD endoglucanase derived from Humicola insolens, DSM 1800 is
 CC described in detail in co-pending Danish patent application No. DK
 CC 1159/90. As a detergent additive, it has a higher degree of specificity
 CC than Bacillus luentis serine protease. The term "higher degree of
 CC specificity" is defined as a protease which conditions degrades human
 CC insulin to fewer components.
 CC
 XX
 SO Sequence 305 AA;
 Query Match 69.9%; Score 1159; DB 13; Length 305;
 Best Local Similarity 67.1%; Pred. No. 5.2e-81;
 Matches 206; Conservative 44; Mismatches 47; Indels 10; Gaps 4;
 YY 1 MRSPVLTITLAALPIVLSAASGSGSTRYWDCKPSCAMPKRAVSOPIVACDANR 60
 DB 1 MRSSPLPSAVVAALPVALALAA--DGRSTRYWDCKPSCGAKKAPVNOPIVFCNANR 58
 YY 61 LSDFNVOGCGN--GGSAYSCADOTPMVAVNDNLAYGFAATSIAGSSSSMCCACALFTFTSG 119
 DB 59 ITDFDAKSGCEGAVAYSCADOTPMVAVNDNLAYGFAATSIAGSNAGMCCACALFTFTSG 118
 YY 120 PVAKTMYVOSTSTGDLGSGNOFDIAMPGGVGVIFNGCSSPFGGLPGAQYGISSRDQCD 179
 DB 119 PVAKKMYVOSTSTGDLGSGNHFDLNIIPGGVGIFDGTCPFGGLPGAQRYGSISSRNECD 178
 YY 180 SFPALKPGCCMRPFQNMADNPTFTFOQVOCPAETVARSGCKRNDSSFP-VFTPPSGG 238
 DB 179 RFPALKPGCYWRPFQNMADNPTFTFOQVOCPAELVARTGCKRNDSSFPVAVQIPSSST 238
 YY 239 NGGTGTPSTAPAGSGQTS-----PGGSGGCTSQKMAQCGGIGFGCTTCVSGTTCOKLN 292
 DB 239 SSPVNPQTSTSTSTSTSSPPVQPTTPSGCTAERMAQCGGNGMGGCTTCVAGSTCTKIN 298
 YY 293 DYYSQCL 299
 DB 299 DWYHQCL 305
 RESULT 11
 AAR37150
 ID AAR37150 standard; Protein; 305 AA.
 XX
 AC AAR37150;
 XX
 DT 25-AUG-1993 (first entry)
 XX
 DE Dye transfer inhibiting compsn. cellulase.
 XX
 KM Detergent; homogeneous endoglucanase component; 43kD cellulase.
 XX
 OS Humicola insolens.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..21
 FT /note= "signal peptide"
 FT Peptide 22..305
 FT /note= "mature peptide"
 XX
 PN EPP540784-A.

XX 12-MAY-1993.
 XX
 XX 06-NOV-1991; 91EP-0202882.
 PE
 XX 06-NOV-1991; 91EP-0202882.
 PR
 XX (PROC) PROCTER & GAMBLE CO.
 PA
 XX Busch A, McCorquodale F;
 PI
 XX WPI; 1993-153868/19.
 DR
 DR N-PSDB; AAQ41732.
 XX
 PT Compsn. for inhibiting dye transfer during fabric washing - contains
 PT pectinase, hydrogen peroxide, substrate and cellulase, esp. endo-
 PT glucanase from Humicola insolens
 XX
 PS Disclosure; Page 18-19; 28pp; English.
 XX
 CC The sequence is that of a cellulase which is characterised in
 CC that the cellulase provides at least 10% removal of immobilised
 CC radioactive labelled carboxymethyl cellulose according to the C14CMC
 CC method at 25 x 10(-6)% by weight of the cellulase protein in the
 CC test solution. It can be used as part of a compsn. for inhibiting
 CC dye transfer.
 CC
 XX
 SO Sequence 305 AA;
 Query Match 69.9%; Score 1159; DB 14; Length 305;
 Best Local Similarity 67.1%; Pred. No. 5.2e-81;
 Matches 206; Conservative 44; Mismatches 47; Indels 10; Gaps 4;
 YY 1 MRSTVLTITLAALPIVLSAASGSGSTRYWDCKPSCAMPKRAVSOPIVACDANR 60
 DB 1 MRSSPLPSAVVAALPVALALAA--DGRSTRYWDCKPSCGAKKAPVNOPIVFCNANR 58
 YY 61 LSDFNVOGCGN--GGSAYSCADOTPMVAVNDNLAYGFAATSIAGSSSSMCCACALFTFTSG 119
 DB 59 ITDFPAKSGCEGAVAYSCADOTPMVAVNDNLAYGFAATSIAGSNAGMCCACALFTFTSG 118
 YY 120 PVAKTMYVOSTSTGDLGSGNOFDIAMPGGVGVIFNGCSSPFGGLPGAQYGISSRDQCD 179
 DB 119 PVAKKMYVOSTSTGDLGSGNHFDLNIIPGGVGIFDGTCPFGGLPGAQRYGSISSRNECD 178
 YY 180 SFPALKPGCCMRPFQNMADNPTFTFOQVOCPAETVARSGCKRNDSSFP-VFTPPSGG 238
 DB 179 RFPALKPGCYWRPFQNMADNPTFTFOQVOCPAELVARTGCKRNDSSFPVAVQIPSSST 238
 YY 239 NGGTGTPSTAPAGSGQTS-----PGGSGGCTSQKMAQCGGIGFGCTTCVSGTTCOKLN 292
 DB 239 SSPVNPQTSTSTSTSTSSPPVQPTTPSGCTAERMAQCGGNGMGGCTTCVAGSTCTKIN 298
 YY 293 DYYSQCL 299
 DB 299 DWYHQCL 305
 RESULT 12
 AAR42063
 ID AAR42063 standard; Protein; 305 AA.
 XX
 AC AAR42063;
 XX
 DT 28-APR-1994 (first entry)
 XX
 DE Endoglucanase enzyme.
 XX
 KM Detergent; fabric; surfactant; softening clay; cellulase;
 KM Humicola insolens; Fusarium oxysporum; endoglucanase.
 XX
 OS Humicola insolens.
 XX

FH	Key	Location/Qualifiers
FT	Peptide	1..21
FT	Protein	/label= sig_peptide 22..305
XX		/label= mat_protein
XX		
XX	AV9211048-A.	
XX		
XX	02-SEP-1993.	
XX		
XX	18-FEB-1992;	92AU-0011048.
XX		
XX	18-FEB-1992;	92AU-0011048.
XX	(PROC) PROCTER & GAMBLE CO.	
XX	Baeck AC, Busch A, Convents AC;	
XX		
DR	WPI: 1993-328419/42.	
XX	N-PSDB; AAQ49941.	
XX		
PT	Detergent compsns., esp. for fabrics - contain surfactant,	
PT	sintering clay and high activity cellulase, partic. from Humicola	
PT	insolens	
PS	Claim 5; Page 54-55; 71pp; English.	
XX		
CC	A new detergent comprises a high activity cellulase in combination	
CC	with a softening clay. The cellulase may be an endoglucanase	
CC	enzyme derived from Humicola insolens (AAQ49941) or Fusarium oxysporum	
CC	(AAQ49942). The combination provides a higher than additive softening	
CC	performance and excellent colour rejuvenation and whiteness	
CC	maintenance for fabrics.	
XX		
SQ	Sequence 305 AA;	
	Query Match 69.9%; Score 1159; DB 14; Length 305;	
	Best Local Similarity 67.1%, Pred. No. 5.2e-81;	
	Matches 206; Conservative 44; Mismatches 47; Indels 10; Gaps	
OY	1 MRSTVLTTLTTLAALPLVASAASGSGSTRYWDCKKPCSCAMPGKAAYSPVYACDANFOR	60
Db	1 MRSSPLLPSAVVAALPVLALAA--DGRSTRYWDCKKPCSGMAKKAPVNGPVSCNNANFOR	58
OY	61 LSDFNVSOGCN-GGSAYSCADQTPPAVNDNLAYGPRAISIASGSESSWCACATLTF7SG	119
Db	59 ITDFEAKSGCEHGCVAYSCADQTPPAVNDNFALGFMAAISIASNSEAGMCACAYELTF7SG	118
OY	120 PVAGKTMYVOASTSGGDLSNOFDIAMPPEGGVGINFGCSOFGGLPGAAYGGISSRDOD	179
Db	119 PVAGKKMYYOASTSGDLAGSNHFDLNIFEGGVGIFEDGCTPFGGLPGAQRXGGISSRNECD	178
OY	180 SEPAPLRGCGOWRFDMFONADNPFTTFDOVOCPAETIVARSCKRRNDSSFP-VFTPPSGG	238
Db	179 RFPDALRKFCGYRFRPMFKADNPSSFSGRYQCPEAEIVARTGCRNRDGNFPAVQIPSSST	238
OY	239 NGGTFTPTSIAGSQT-----FGGSGGCSOKMAOCGGIGFSGCTCYSGTTQQIKIN	292
Db	239 SSPVNQPTSTSTSTSTSYSSPPVQFTTPBSGCTAERMAQCGGMWSGCTCVAGSTCTKIN	298
OY	293 DYSSQCL 299	
Db	299 DMYHQCL 305	
	RESULT 13	
	AAy67388	
	ID AAY67388 standard; Protein; 305 AA.	
XX	AA67388;	
XX		
XT	25-APR-2000 (first entry)	

DE	H. insolens endoglucanase enzyme protein sequence.
XX	
KM	Endoglucanase; cellulase enzyme; detergent composition; anion surfactant;
KW	degradation protection; liquid detergent; long term stability.
XX	
OS	Humicola insolens.
PN	EP633311-A1.
XX	
PD	11-JAN-1995.
PF	28-JUN-1993; 93EP-0870122.
XX	
PR	28-JUN-1993; 93EP-0870122.
XX	
PA	(PROC) PROCTER & GAMBLE CO.
XX	
P1	Herbots IMAJ, Jansen MP;
XX	
DR	WPI; 1995-038508/06.
DR	N-PSDB; AAZ60178.
XX	
PT	Liq.detergent compsns. - contg. hydrophobic amine(s) for cellulase
XX	stabilisation
PS	Claim 11; Page 18-19; 28pp; English.
XX	
CC	This sequence represents the Humicola insolens endoglucanase enzyme
CC	protein sequence. The enzyme is a cellulase enzyme used in the liquid
CC	detergent composition of the invention. The detergent comprises an anion
CC	surfactant, a cellulase enzyme and a cellulase stabilising amount of an
CC	amide of the formula R1R2RN; where R1 and R2 are independently H or a
CC	C2-C9 alkyl chain, and R3 is a C2-C9 alkyl chain or cyclohexyl or
CC	cyclopentyl or cycloheptyl. The amine prevents the cellulase enzyme from
CC	becoming degraded. The composition is used as a liquid detergent, and has
CC	long term stability. It is applicable to the protection of a cellulase
CC	and can be used in the presence of proteases.
XX	
SQ	Sequence 305 AA:
Query Match	69.9%; Score 1159; DB 16; Length 305;
Best Local Similarity	67.1%; Pred. No. 5.2e-81;
Matches 206; Conservative 44; Mismatches 47; Indels 10; Gaps 4,	
OY	1 MRSTVLRTTLAALPLVASAASGSGSTRYYDCKKPCSCAMPKKAAYSPPVACANFOR 60
Db	1 MRSSELPLESAVVAAAPLVIALA--DGSRTRYWDCKKPCSGAKKAPVNVPVSCANFQR 58
OY	61 LSDFNWGSCN-GGSAYSCADOTPVAVNDNLVGYFAATSIAGSSESSWCACAYALTFTSG 119
Db	59 ITDFDAKSGCGGCVAYASCADOTPVAVNDLFLGRAITSIAGSNEMGWCACTELFTSG 118
OY	120 PVAGTMMVOSTSTGGDLGNSQFDIAMPGGVGITENGSSQFGGLPGAOLYGISRDQCD 179
Db	119 PVAAGKKMVQOSTGTGDLSNHFDLINIPGGVGIFEGCIPQFGGLPGGRYGGISRNEDC 178
OY	180 SPPARLKPGCOMRPWFOMANDPTFFEOVOCPAEIVASGGKRNDSSF-P-VFRPPSGG 238
Db	179 RFDDALKPECTRYRFDMFKADNPSTSFROYQCPALVIATTCRRNDGNGFPNAVQIPSSST 238
OY	239 NGTGTPMTSTABSGQTS-----PQGGSGCTSQKWAOCGGIGTSGCTTVCSGTTCQKLN 292
Db	239 SSPVNOQTSTSTSTSTSSPVPQPTTPSGCTAERWACGGGMWSGCTTCVAGSTCTCKIN 298
OY	293 DYYSOCL 299
Db	299 DMYHOCL 305
RESULT 14	
ID	AAW01502
XX	AAW01502 standard: protein; 305 AA.

AC	AAW01502;
XX	
DT	26-FEB-1997 (first entry)
XX	
DE	43 kD endoglucanase, EG V.
XX	
XX	Detergent composition; cellulase; retaining-type activity; catalytic
KM	activity; cellobiose; particulate soil removal; colour clarification;
KM	cleaning; cellulose-containing fabric; cellulohydrolase; endoglucanase;
XX	
OS	<i>Humicola insolens</i> , DSM 1800.
XX	

FH	Key	location/qualifiers
FT	peptide	1..21
FT		/label= signal_peptide
FT	protein	22..305
FT		/label= mature_protein
VY		

PN W09502675-A1.
XX

PD 26-JAN-1995

PF 07-JUL-1994; 94WO-DK00280.

PR 11-OCT-1993; 93DK-0001135.

PR 12-JUL-1993; 93EP-0870131.

PA (NOVO) NOVO-NORDISK A.S.

PA
XX
(PROC) PROCTER & GAMBLE CO.
XX

PL converts AC, Jeffreys B, Schueleln M, Tikhomirov DF,

WPL: 1995-06/325/09.

PM Detergent compn. contg. two cellulase components - the first

CC Determining compositions comprising: 1) a first cellulase component having
CC retaining-type activity, pref. having a catalytic activity on cellobiose
CC at pH 8.5 corresponding to kcat of at least 0.01 s⁻¹, and capable of
CC particulate soil removal; and 2) a second cellulase component having
CC multiple domains comprising at least one non-catalytic domain attached to
CC a catalytic domain, pref. having a catalytic activity on Red Avicel 7.5
CC per 1 mg of cellulase protein higher than 10⁻⁶ IU and being capable of
CC colour clarification, where at least one of the cellulase components is a
CC single (recombinant) component, are useful for cleaning and colour
CC clarification of cellulose-containing fabrics. The second cellulase
CC component can be an endoglucanase which is immunoreactive with an
CC antibody raised against a highly purified ~43 kD endoglucanase
CC derived from Humicola insolens, DSM 1800, and is esp. the present
CC sequence, designated Eg V.

50 Sequence 305 AA;

Query Match	Score	DB	Length
69.98;	1159;	16;	305;

Matches	206;	Conservative	44;	Mismatches	47;	Indels	10;	Gaps	4
---------	------	--------------	-----	------------	-----	--------	-----	------	---

[illegible]

QY	180	SEFPALKGCRCRPFDMFONADNPETFEQYOCPAEYARSGCKRNDSSPF-VETPPSGG	238
Db	179	RPFDALKGCYWRDWMFNADNPFSFSPFOYCCPAELYARTGCRNDGDNFPFAVOIPSSST	238
QY	239	NGGTCPTPTSTAPGSGQTS-----PGGSGGCTSOCKMOCGGIGRSGCTTCVSGTTCQCKLN	292
	:		
Db	239	SSPYNQPTSTSTTSTTSTSSPPVQPTTPSGCTAARMACGGCMGSGCTTCVAGSGTCTKIN	298
QY	293	DYSSQCL	299
Db	299	DWYHQCL	305

RESULT 15

ID AAR88471 standard; Protein; 305 AA.

AC AAR88471

DT 03-APR-1996 (first entry)

Alkaline endoglucanase Carezyme.

Alkaline endoglucanase; carezyme; cellulase; host cell;

2000

XXXXXX

1.21 Peptide

XX
1
2
3
4
5
6
7
8
9
10

XX 11-TAN-100C
DD

XX
XX
TIME TO GO
CLOCK
CLOCK
CLOCK

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

PK 30-JUN-1994; 9405-0269449.

PA (NOVO) NOVO NORDISK BIOTECH INC.

PL Moyer DL, Royer JC, Shuster JR, Yoder W;
 YV

DK WPL; 1996-077498/
DE N-DCDD; 22710182

XX
DE
NOV 3 4 5 6 7 8 9 10 11 12

cell - used to produce heterologous p

The Humicola insolens DSM 1800 alkaline endoglucanase Carezyme (AAR8471), a derivative of a 43 kDa cellulase, is expressed in non-toxic, non-toxicogenic, non-pathogenic, protease-deficient recombinant Fusarium graminearum ATCC 20334 carrying plasmid pDM51. This plasmid includes the Carezyme gene (Arl10182) operably linked to promoter and terminator (Arl10184-85) sequences of the Fusarium oxysporum trypsin-like protease sp87 gene. Recombinant enzyme is produced at levels of 6.0 g/l host cell; over 90% of secreted protein is Carezyme.

Sequence 305 AA;

Query Match	69.98;	Score 1159;	DB 17;	Length 305;
-------------	--------	-------------	--------	-------------

Matches	206;	Conservative	44;	Mismatches	47;	Indels	10;	Gaps	4;
---------	------	--------------	-----	------------	-----	--------	-----	------	----

QY 1 MRSTPVLRTTLAALPLVASAASGSGSQTIRYWDCKKPSGAWPGKAASQPVYACDANFQR 60
 |||::| : |||::| | : ||| ||||| ||||| | |||::|: |||||
 Db 1 MRSSPLPESAVVAALPYLALAA--DGRSTRYWDCKKPSGAWKAKAPVNDPVESCANNFQR 58

```

OY 61 LSDFNVOGCGN-GGSAYSCADOTPMAVNDNLAYGEPAATSIAGSESSWCCACATLFTTSG 119
Db 59 ITDFDAKSGCEPGVAYSCADOTPMAVNDNDFALGFAATSIAGSNEAGWCCACACELFTTSG 118
OY 120 PVAGKTMYVOSTSTGTGDLGSGNOFDITAMPGGGVGIFENGCSQFGGLPGAQYGGISRPDCD 179
Db 119 PVAGKRMVYOSTSTGTGDLGSHHFDLNPGGGVGIFDGCTPQFGGLPQRTGIGISSRNECD 178
OY 180 SEPAPLKPQCQWRPFQCNADNPTFTFOOVQCPAEIIVARSCKRNDSSFP-VFTPPSGG 238
Db 179 RPPDALKPEGCYWRPFQCNADNPTFTFOOVQCPAEIIVARSCKRNDSSFP-VFTPPSGG 238
OY 239 NGGTGTPISTAPGSGOTS-----PGGSGCTSQKMAQCGGIGFSGCTTCVSGTTCOKLN 292
Db 239 SSPVNOPTSTSTSTSSPPVQPTTFSGCTAERMAQCGGNGMSGCTTCVAGSTCTKIN 298
OY 293 DYYSOCL 299
Db 299 DMYHOCL 305

```

Search completed: February 24, 2003, 11:51:32
 Job time : 86 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 24, 2003, 11:38:03 ; Search time 14 Seconds

(without alignments)
628,390 Million cell updates/sec

Title: US-10-007-521-12

Perfect score: 1659
Sequence: 1 MRSPTVLRRTTLAALPLVAS.....TTCVSGTTCOKLNDYSSQCL 299

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/prodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/prodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/prodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/prodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/prodata/1/1aa/PCTUS.COMB.pep:*
6: /cgn2_6/prodata/1/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1659	100.0	299	2	US-08-872-437-2
2	1659	100.0	299	3	US-08-651-136C-12
3	1659	100.0	299	4	US-09-229-911A-6
4	1178	71.0	305	4	US-09-230-222-1
5	1159	69.9	305	1	US-08-090-013-2
6	1159	69.9	305	1	US-08-081-328-2
7	1159	69.9	305	1	US-08-232-249-2
8	1159	69.9	305	2	US-08-921-426-8
9	1159	69.9	305	2	US-08-833-642A-2
10	1159	69.9	305	2	US-08-140-008A-4
11	1159	69.9	305	2	US-08-836-340-1
12	1159	69.9	305	2	US-08-389-423-2
13	1159	69.9	305	3	US-08-816-915-8
14	1159	69.9	305	4	US-09-189-060B-56
15	1159	69.9	305	4	US-09-230-665-2
16	1159	69.9	305	4	US-09-189-028-2
17	1159	69.9	305	5	PCT-US95-07743-8
18	1142.5	67.9	306	4	US-09-189-060B-66
19	1137	67.9	286	4	US-09-254-733-3
20	1125	67.8	285	4	US-09-230-225B-6
21	1112	67.0	284	1	US-08-411-777-10
22	1112	67.0	284	3	US-09-057-088-10
23	1107.5	66.8	306	4	US-09-189-060B-68
24	1077.5	64.9	308	4	US-09-189-060B-70
25	1076.5	64.9	307	4	US-09-189-060B-72
26	1045	63.0	307	4	US-09-189-060B-74
27	1039	62.6	235	4	US-09-329-350-31

28	1021	61.5	308	3	US-08-651-136C-6	Sequence 6, Appl
29	1021	61.5	308	4	US-09-229-911A-6	Sequence 6, Appl
30	994.5	59.9	297	3	US-08-651-136C-4	Sequence 4, Appl
31	994.5	59.9	297	4	US-09-229-911A-4	Sequence 4, Appl
32	968.5	58.4	234	4	US-09-230-665-6	Sequence 6, Appl
33	962	58.0	310	3	US-08-651-136C-22	Sequence 22, Appl
34	962	58.0	310	4	US-09-229-911A-22	Sequence 22, Appl
35	951	57.3	295	3	US-08-651-136C-8	Sequence 8, Appl
36	951	57.3	295	4	US-09-229-911A-8	Sequence 8, Appl
37	939.5	56.6	376	1	US-08-090-013-4	Sequence 4, Appl
38	939.5	56.6	376	1	US-08-081-328-4	Sequence 4, Appl
39	939.5	56.6	376	1	US-08-232-249-4	Sequence 4, Appl
40	939.5	56.6	376	2	US-08-833-642A-4	Sequence 4, Appl
41	939.5	56.6	376	2	US-08-389-423-4	Sequence 4, Appl
42	939.5	56.6	376	4	US-09-189-028-4	Sequence 4, Appl
43	934.5	56.3	376	4	US-09-230-665-4	Sequence 4, Appl
44	923.5	55.7	294	3	US-08-651-136C-24	Sequence 24, Appl
45	923.5	55.7	294	4	US-09-229-911A-24	Sequence 24, Appl

ALIGNMENTS

```
RESULT 1
US-08-872-437-2
; Sequence 2, Application US/08872437
; Patent No. 5958082
; GENERAL INFORMATION:
; APPLICANT: Lund, Henrik
; APPLICANT: Kalum, Lisbeth
; TITLE OF INVENTION: Garments With Considerable Variation In
; TITLE OF INVENTION: Abrasion Level
; FILE REFERENCE: 4888, 200-US
; CURRENT APPLICATION NUMBER: US/08/872,437
; CURRENT FILING DATE: 1997-06-10
; EARLIER APPLICATION NUMBER: 1276/96
; EARLIER FILING DATE: 1996-11-13
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Thielavia terrestris
US-08-872-437-2

Query Match      100.0%; Score 1659; DB 2; Length 299;
Best Local Similarity 100.0%; Pred. No. 1.6e-127;
Matches 299; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRSPTVLRRTTLAALPLVASAASGOSTRYWDCCKPSCAMPKAAVSQPVYACDANFOR 60
Db 1 MRSPTVLRRTTLAALPLVASAASGOSTRYWDCCKPSCAMPKAAVSQPVYACDANFOR 60

QY 61 LSDNVVSGCGSGAYSACADQTPMAVNDNLAYGFAATSIAGSSSSWCCACALFTTSGP 120
Db 61 LSDNVVSGCGSGAYSACADQTPMAVNDNLAYGFAATSIAGSSSSWCCACALFTTSGP 120

QY 121 VAGTMYVVGSTSTGDLGNSNFDIAMPBGVGINCGSSQSGCLPGAQYIGTISRDDQDS 180
Db 121 VAGTMYVVGSTSTGDLGNSNFDIAMPBGVGINCGSSQSGCLPGAQYIGTISRDDQDS 180

QY 181 FPAALKPGCCQRFPMFQVADNPFTFOQVCPAETVARSGCKRNDSSFPVFTPPSGNG 240
Db 181 FPAALKPGCCQRFPMFQVADNPFTFOQVCPAETVARSGCKRNDSSFPVFTPPSGNG 240

QY 241 GTGPTSTAPSGSGTSPGSGGCTSQKWAQCGGIGFSCTTCVSGTTQCKLNDYSSQCL 299
Db 241 GTGPTSTAPSGSGTSPGSGGCTSQKWAQCGGIGFSCTTCVSGTTQCKLNDYSSQCL 299

RESULT 2
US-08-651-136C-12
; Sequence 12, Application US/08651136C
```

```

; Patent No. 6001639
; GENERAL INFORMATION:
; APPLICANT: Schulein, Martin
; APPLICANT: Andersen, Lene N.
; APPLICANT: Lassen, Soren F.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Lange, Lene
; APPLICANT: Nielsen, Ruby I.
; APPLICANT: Ihara, Michiko
; APPLICANT: Takagi, Shinobu
; TITLE OF INVENTION: No. 6001639e1 Endoglucanases
; NUMBER OF SEQUENCES: 109
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6001639o No. 6001639disk of No. 6001639th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/651,136C
; FILING DATE: 21-MAY-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambitis, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 4366.200-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 299 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-651-136C-12

```

```

Query Match      100.0%; Score 1659; DB 3; Length 299;
Best local Similarity 100.0%; Pred. No. 1.6e-127;
Matches 299; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 MRSTPVLRTTLAALPLVSAASGSGSTRYWDCKPSCAMPKRAAVSOPVYACDANFOR 60
    |||||||
DB 1 MRSTPVLRTTLAALPLVSAASGSGSTRYWDCKPSCAMPKRAAVSOPVYACDANFOR 60
QY 61 LSDFNVOGCGNGSAYSCAODTPWAVNDNLAYGFAATSIAGSSSSWCACALFTFTSGP 120
    |||||||
DB 61 LSDFNVOGCGNGSAYSCAODTPWAVNDNLAYGFAATSIAGSSSSWCACALFTFTSGP 120
QY 121 VAGKTMVOSTSTGDLGSGNFDIAMPGGVGIIFNGCSSQFGGLPGAQYGGISSRDQDS 180
    |||||||
DB 121 VAGKTMVOSTSTGDLGSGNFDIAMPGGVGIIFNGCSSQFGGLPGAQYGGISSRDQDS 180
QY 121 VAKTMYVOSTSTGDLGSGNFDIAMPGGVGIIFNGCSSQFGGLPGAQYGGISSRDQDS 180
    |||||||
DB 121 VAKTMYVOSTSTGDLGSGNFDIAMPGGVGIIFNGCSSQFGGLPGAQYGGISSRDQDS 180
QY 181 FPAPLPGCGQWRDWFONADNPTFTFOQVCPAETIVARSGCKRNDSSFPVFTPPSGGNG 240
    |||||||
DB 181 FPAPLPGCGQWRDWFONADNPTFTFOQVCPAETIVARSGCKRNDSSFPVFTPPSGGNG 240
QY 241 GTCTPTSTAGSGQTSPPGGSGCTSQKWAQCGGIGFSGCTTCVSGTTCOKLNDYISQCL 299
    |||||||
DB 241 GTCTPTSTAGSGQTSPPGGSGCTSQKWAQCGGIGFSGCTTCVSGTTCOKLNDYISQCL 299

```

```

RESULT 3
US-09-229-911A-12
; Sequence 12, Application US/09229911A
; Patent No. 6387690
; GENERAL INFORMATION:

```

```

; APPLICANT: Schulein, Martin
; APPLICANT: Andersen, Lene N.
; APPLICANT: Lassen, Soren F.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Lange, Lene
; APPLICANT: Nielsen, Ruby I.
; APPLICANT: Ihara, Michiko
; APPLICANT: Takagi, Shinobu
; TITLE OF INVENTION: No. 6387690e1 Endoglucanases
; NUMBER OF SEQUENCES: 109
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6387690o No. 6387690disk of No. 6387690th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/229,911A
; FILING DATE: 13-Jan-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/651,136
; FILING DATE: 21-MAY-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambitis, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 4366.200-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 299 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 12:
; US-09-229-911A-12

```

```

Query Match      100.0%; Score 1659; DB 4; Length 299;
Best local Similarity 100.0%; Pred. No. 1.6e-127;
Matches 299; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 MRSTPVLRTTLAALPLVSAASGSGSTRYWDCKPSCAMPKRAAVSOPVYACDANFOR 60
    |||||||
DB 1 MRSTPVLRTTLAALPLVSAASGSGSTRYWDCKPSCAMPKRAAVSOPVYACDANFOR 60
QY 61 LSDFNVOGCGNGSAYSCAODTPWAVNDNLAYGFAATSIAGSSSSWCACALFTFTSGP 120
    |||||||
DB 61 LSDFNVOGCGNGSAYSCAODTPWAVNDNLAYGFAATSIAGSSSSWCACALFTFTSGP 120
QY 121 VAGKTMVOSTSTGDLGSGNFDIAMPGGVGIIFNGCSSQFGGLPGAQYGGISSRDQDS 180
    |||||||
DB 121 VAGKTMVOSTSTGDLGSGNFDIAMPGGVGIIFNGCSSQFGGLPGAQYGGISSRDQDS 180
QY 121 VAKTMYVOSTSTGDLGSGNFDIAMPGGVGIIFNGCSSQFGGLPGAQYGGISSRDQDS 180
    |||||||
DB 121 VAKTMYVOSTSTGDLGSGNFDIAMPGGVGIIFNGCSSQFGGLPGAQYGGISSRDQDS 180
QY 181 FPAPLPGCGQWRDWFONADNPTFTFOQVCPAETIVARSGCKRNDSSFPVFTPPSGGNG 240
    |||||||
DB 181 FPAPLPGCGQWRDWFONADNPTFTFOQVCPAETIVARSGCKRNDSSFPVFTPPSGGNG 240
QY 241 GTCTPTSTAGSGQTSPPGGSGCTSQKWAQCGGIGFSGCTTCVSGTTCOKLNDYISQCL 299
    |||||||
DB 241 GTCTPTSTAGSGQTSPPGGSGCTSQKWAQCGGIGFSGCTTCVSGTTCOKLNDYISQCL 299

```

```

RESULT 4
US-09-230-222-1
; Sequence 1, Application US/09230222A

```


COMPUTER: IBM PC compatible

```

COMPIER READABLE FORM:
MEDIUM TYPE: Diskette

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/081,328
FILING DATE: 11/19/93
CLASSIFICATION: 252
ATTORNEY/AGENT INFORMATION:
NAME: ZERRY, KIM WILLIAM
REGISTRATION NUMBER: 32,323
REFERENCE/DOCKET NUMBER: CM356M
TELECOMMUNICATION INFORMATION:
TELEPHONE: 513-627-2885
TELEFAX: 513-627-0318
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 305 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-081-328-2

Query Match 69.9%; Score 1159; DB 1; Length 305;
Best Local Similarity 67.1%; Pred. No. 7, 1e-87;
Matches 206; Conservative 44; Mismatches 47; Indels 10; Gaps 4;

QY 1 MSTPVLRTTLAALPLVSAASGSGSTRYWDCKRSCAMPKAAVSPVYACDANFOR 60
DB 1 MRSPLPLPSVAVALPVLALAA--DGRSTRYWDCKRSCAMPKAAVSPVYACDANFOR 58
QY 61 LSFDFVNSGCN--GGSAYSCADQTPMAVNDNLAYGFAATSTAGSESSWCACALFTFTSG 119
DB 59 IYDFDAKSGCEPGGVAYSCADQTPMAVNDNLAYGFAATSTAGSESSWCACALFTFTSG 118
QY 120 PVAGKTMVYSTSTGDLGSGNDFDIAMPGGGVGIFNGCSSQFGLPQAGYGGTSSRQCD 179
DB 119 PVAGKTMVYSTSTGDLGSGNDFDIAMPGGGVGIFNGCSSQFGLPQAGYGGTSSRQCD 178
QY 180 SPPAPLPCGCMRFDFMADNPTFTFOYQCPAEIVARSGCKRNDSSPP-VFTPPSGG 238
DB 179 RFPDALKPGCYRMFDFMADNPTFTFOYQCPAEIVARSGCKRNDSSPP-VFTPPSGG 238
QY 239 NGGTGPTSTAPSGGOTS-----PGGSGCTSQKMAQCGGIGFSGCTTVSGTTCOKLN 292
DB 239 SSPVNPPTSTSTSTSSPPVQPTTPSGCTAERMAQCGGNGMWSGCTTVAGSTCTKIN 298
QY 293 DYYSOCL 299
DB 299 DMYHOCL 305

RESULT 7
US-08-232-249-2
Sequence 2, Application US/08232249
Patent No. 5610129
GENERAL INFORMATION:
APPLICANT: MCCORODALE, FINLAY (NMN)
APPLICANT: BUSCH, ALFRED (NMN)
TITLE OF INVENTION: DYE TRANSFER INHIBITING COMPOSITIONS
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESSES:
ADDRESSEE: THE PROCTER & GAMBLE COMPANY
STREET: 5299 SPRING GROVE AVENUE
CITY: CINCINNATI
STATE: OHIO
COUNTRY: USA
ZIP: 45217
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/232,249
FILING DATE: 05-MAY-1994
CLASSIFICATION: 510
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 91202882.6
FILING DATE: 06-NOV-91
ATTORNEY/AGENT INFORMATION:
NAME: ALLEN, GEORGE W.
REGISTRATION NUMBER: 26,143
REFERENCE/DOCKET NUMBER: CM-395
TELECOMMUNICATION INFORMATION:
TELEPHONE: 513-627-5946
TELEFAX: 513-627-8118
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 305 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-232-249-2

Query Match 69.9%; Score 1159; DB 1; Length 305;
Best Local Similarity 67.1%; Pred. No. 7, 1e-87;
Matches 206; Conservative 44; Mismatches 47; Indels 10; Gaps 4;

QY 1 MSTPVLRTTLAALPLVSAASGSGSTRYWDCKRSCAMPKAAVSPVYACDANFOR 60
DB 1 MRSPLPLPSVAVALPVLALAA--DGRSTRYWDCKRSCAMPKAAVSPVYACDANFOR 58
QY 61 LSFDFVNSGCN--GGSAYSCADQTPMAVNDNLAYGFAATSTAGSESSWCACALFTFTSG 119
DB 59 IYDFDAKSGCEPGGVAYSCADQTPMAVNDNLAYGFAATSTAGSESSWCACALFTFTSG 118
QY 120 PVAGKTMVYSTSTGDLGSGNDFDIAMPGGGVGIFNGCSSQFGLPQAGYGGTSSRQCD 179
DB 119 PVAGKTMVYSTSTGDLGSGNDFDIAMPGGGVGIFNGCSSQFGLPQAGYGGTSSRQCD 178
QY 180 SPPAPLPCGCMRFDFMADNPTFTFOYQCPAEIVARSGCKRNDSSPP-VFTPPSGG 238
DB 179 RFPDALKPGCYRMFDFMADNPTFTFOYQCPAEIVARSGCKRNDSSPP-VFTPPSGG 238
QY 239 NGGTGPTSTAPSGGOTS-----PGGSGCTSQKMAQCGGIGFSGCTTVSGTTCOKLN 292
DB 239 SSPVNPPTSTSTSTSSPPVQPTTPSGCTAERMAQCGGNGMWSGCTTVAGSTCTKIN 298
QY 293 DYYSOCL 299
DB 299 DMYHOCL 305

RESULT 8
US-08-921-426-8
Sequence 8, Application US/08921426
Patent No. 5837847
GENERAL INFORMATION:
APPLICANT: Royer, John C
APPLICANT: Moyer, Donna L
APPLICANT: Yoder, Wendy T
APPLICANT: Spuster, Jeffrey R
TITLE OF INVENTION: NON-TOXIC, NON-ROXIGENIC, NON-PATHOGENIC
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESSES:
ADDRESSEE: No. 5837847 of No. 5837847 of No. 5837847th America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

```

1      SOFTWARE: PatentIn Release #1.0, Version #1.30
2      CURRENT APPLICATION DATA:
3      APPLICATION NUMBER: US/08/921,426
4      FILING DATE: 29-AUG-1997
5      CLASSIFICATION: 435
6      PRIOR APPLICATION DATA:
7      APPLICATION NUMBER: US 08/456,433
8      FILING DATE: 01-JUN-1995
9      APPLICATION NUMBER: US 08/404,678
10     FILING DATE: 15-MAR-1995
11     ATTORNEY/AGENT INFORMATION:
12     NAME: Agtis Dr., Cheryl H.
13     REGISTRATION NUMBER: 34,086
14     REFERENCE/DOCKET NUMBER: 4216, 010-US
15     TELECOMMUNICATION INFORMATION:
16     TELEPHONE: 212-867-0123
17     TELEFAX: 212-878-9655
18     INFORMATION FOR SEQ ID NO: 8:
19     SEQUENCE CHARACTERISTICS:
20     LENGTH: 305 amino acids
21     TYPE: amino acid
22     TOPOLOGY: linear
23     MOLECULE TYPE: protein
24     US-08-921-426-8

```

Query Match	69.9%	Score 1159	DB 2	Length 305
Best Local Similarity	67.1%	Pred. No. 7.1e-87		
Matches 206	Conservative 44	Mismatches 47	Indels 10	Gaps 4

[illegible]

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

```

1 MEDIUM TYPE: Diskette, 3.5 Inch
2 COMPUTER: IBM PC Compatible
3 OPERATING SYSTEM: MS-DOS
4 SOFTWARE: WordPerfect 6.1
5
6 CURRENT APPLICATION DATA:
7 APPLICATION NUMBER: US/08/833, 642A-2
8 FILING DATE: April 8, 1997
9
10 ATTORNEY/AGENT INFORMATION:
11 NAME: Zurich, J. A.
12 REGISTRATION NUMBER: P42,251
13 REFERENCE/DOCKET NUMBER: CM551C
14 TELECOMMUNICATION INFORMATION:
15 TELEPHONE: (313) 977-8377
16 TELEFAX: (513) 977-8141
17
18 INFORMATION FOR SEQ ID NO: 2:
19 SEQUENCE CHARACTERISTICS:
20 LENGTH: 305 amino acids
21 TYPE: amino acid
22 TOPOLOGY: linear
23
24 MOLECULE TYPE: protein
25
26 US-08-833-642A-2

```

Query Match	69.9%	Score 1159;	DB 2;	Length 305;
Best Local Similarity	67.1%	Pred. No. 7.1e-87;		
Matches 206;	Conservative 44;	Mismatches 47;	Indels 10;	Gaps 4;

[illegible]

RESULT 10
US-08-140-008A-4
Sequence 4, Application US/08140008A
Patent No. 5914306
GENERAL INFORMATION:
APPLICANT: SVENDSEN, Allan
APPLICANT: VON DER OSTEN, Claus
APPLICANT: CLAUSEN, Ib Groch
APPLICANT: PATIKAR, Shamkant Anant
APPLICANT: BORCH, Kim
TITLE OF INVENTION: STABILIZED ENZYMES
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5914306o No. 5914306disk of No. 5914306th America
STREET: 405 Lexington Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10114
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

NAME: Lambitis, Elias J.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 3469.214-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 305 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-389-423-2

Query Match 69.9%; Score 1159; DB 2; Length 305;
Best Local Similarity 67.1%; Pred. No. 7.1e-87;
Matches 206; Conservative 44; Mismatches 47; Indels 10; Gaps 4;

QY 1 MRSTPVLRITLAAALPLVASAASGSGSTRYWDCKPSCAMPKAAVSQPVYACDANFOR 60
DB 1 MRSSPLLPASVVALPVALALAA--DGRSTRYWDCKPSCAMPKAAVNPQVFSKANFOR 58
QY 61 LSDFNVOGCGN--GSAVSCADQTPMAVNDLAFATSTAGSESSWCCACVALFTTSG 119
DB 59 ITDFDAKSGCEPGGVAVSCADQTPMAVNDLAFATSTAGSESSWCCACVALFTTSG 118
QY 120 PVAGKTMVVOSTSTGDLGNSQFDIAMPGGVGIFNGCSQFGLPGAQYGGISSRQCD 179
DB 119 PVAGKTMVVOSTSTGDLGNSQFDIAMPGGVGIFNGCSQFGLPGAQYGGISSRQCD 178
QY 180 SPFAPLPGCGMRFDFONADNFTFFQOYQCPAEIYARSCKRNDSSFP-VFTPPSGG 238
DB 179 RFPDALPKPCYWFDFMKNADNPFSEFRQYQCPAEIYARFCRRNDGPNPAVQIPSSST 238
QY 239 NGCTGPTSTAPSGQTS-----PGGSGCTSQKNAOCGGIGFSGCTTCVSGTTCKLN 292
DB 239 SSPVNPPTSTSTSTSTSSPPVQPTTPSGCTAERMAOCGGNGSGCTTCVAGSTCKIN 298
QY 293 DYYSOCL 299
DB 299 DWYHOCL 305

RESULT 13
US-08-816-915-8
Sequence 8, Application US/08816915
Patent No. 6060305
GENERAL INFORMATION:
APPLICANT: Royer, John C
APPLICANT: Moyer, Donna L
APPLICANT: Yoder, Wendy T
APPLICANT: Shuster, Jeffrey R
TITLE OF INVENTION: NON-TOXIC, NON-TOXIGENIC, NON-PATHOGENIC
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 60603050 No. 6060305disk of No. 6060305th America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10174-6401

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/816.915
FILING DATE: 13-MAR-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Agilis Dr., Cheryl H.

REGISTRATION NUMBER: 34,086
REFERENCE/DOCKET NUMBER: 4216.240-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 305 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-816-915-8

Query Match 69.9%; Score 1159; DB 3; Length 305;
Best Local Similarity 67.1%; Pred. No. 7.1e-87;
Matches 206; Conservative 44; Mismatches 47; Indels 10; Gaps 4;

QY 1 MRSTPVLRITLAAALPLVASAASGSGSTRYWDCKPSCAMPKAAVSQPVYACDANFOR 60
DB 1 MRSSPLLPASVVALPVALALAA--DGRSTRYWDCKPSCAMPKAAVNPQVFSKANFOR 58
QY 61 LSDFNVOGCGN--GSAVSCADQTPMAVNDLAFATSTAGSESSWCCACVALFTTSG 119
DB 59 ITDFDAKSGCEPGGVAVSCADQTPMAVNDLAFATSTAGSESSWCCACVALFTTSG 118
QY 120 PVAGKTMVVOSTSTGDLGNSQFDIAMPGGVGIFNGCSQFGLPGAQYGGISSRQCD 179
DB 119 PVAGKTMVVOSTSTGDLGNSQFDIAMPGGVGIFNGCSQFGLPGAQYGGISSRQCD 178
QY 180 SPFAPLPGCGMRFDFONADNFTFFQOYQCPAEIYARSCKRNDSSFP-VFTPPSGG 238
DB 179 RFPDALPKPCYWFDFMKNADNPFSEFRQYQCPAEIYARFCRRNDGPNPAVQIPSSST 238
QY 239 NGCTGPTSTAPSGQTS-----PGGSGCTSQKNAOCGGIGFSGCTTCVSGTTCKLN 292
DB 239 SSPVNPPTSTSTSTSTSSPPVQPTTPSGCTAERMAOCGGNGSGCTTCVAGSTCKIN 298
QY 293 DYYSOCL 299
DB 299 DWYHOCL 305

RESULT 14
US-09-189-060B-56
Sequence 56, Application US/09189060B
Patent No. 6270968
GENERAL INFORMATION:
APPLICANT: Dalboge, Henrik
APPLICANT: Sandal, Thomas
APPLICANT: Kaupinen, Markus
APPLICANT: Borge, Diderichsen
TITLE OF INVENTION: Method Of Providing No. 6270968e1 DNA Sequences
FILE REFERENCE: 4772.204-US
CURRENT APPLICATION NUMBER: US/09/189,060B
PRIOR FILING DATE: 1998-11-10
PRIOR APPLICATION NUMBER: PCT/DK97/00216
NUMBER OF SEQ ID NOS: 74
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 56
LENGTH: 305
TYPE: PRP
ORGANISM: Humicola insolens
FEATURE:
NAME/KEY: STGNAL
LOCATION: (1)....(21)
US-09-189-060B-56

Query Match 69.9%; Score 1159; DB 4; Length 305;
Best Local Similarity 67.1%; Pred. No. 7.1e-87;
Matches 206; Conservative 44; Mismatches 47; Indels 10; Gaps 4;

QY 1 MRSTPVLRITLAAALPLVASAASGSGSTRYWDCKPSCAMPKAAVSQPVYACDANFOR 60

Search completed: February 24, 2003, 11:43:05
Job time : 15 secs

```
Db 1 MRSSPLLSAVVAALPVLALAA--DGRSTRYWDCCPKSCGMAKAPVNOVFSCNANFOR 58
QY 61 LSDFNVOGSCN-GGSAYSCADQTPMAVNDNLAYGFAATSIAGSESSWCACAYALFTSG 119
Db 59 ITDFDAKSGCEPGGVASCADQTPMAVNDNDFALGFAATSIAGSNAGWCACAYELFTSG 118
QY 120 PVAGKTMVVOSTSTGGDLGSDNOFDIAMPGGVGIFNGCSSOFGGLPGAQYGGISRPDCD 179
Db 119 PVAGKTMVVOSTSTGGDLGSDNHFDLNPGGVGIFDGCPTQFGGLPGAQYGGISRNEDC 178
QY 180 SFPAPLKPCCOMRPFDMFOVADNPTFTFOOVCPAEIYARSGCKRNDSSFP-VFTPPSGG 238
Db 179 RPPDALKEGCIYRPFDMFKANDNPFSEFROYOCPAEIYARTGCRNDGNFPAVOIPSSST 238
QY 239 NGGTGPTSTAPGSGQTS-----PGGSGCTSQKMAQCGGIGPSGCTTCVSGTTCOKLN 292
Db 239 SSPVNOPTSTSTSTSTSSPPVQPTTPSGCTAERWAQCGGNGMSGCTTCVAGSTCTKIN 298
QY 293 DYYSQCL 299
Db 299 DWYHOCCL 305
```

```
RESULT 15
US-09-230-665-2
; Sequence 2, Application US/09230665
; Patent No. 6322595
; GENERAL INFORMATION:
; APPLICANT: Boyer, Stanton L
; TITLE OF INVENTION: Detergent Composition Comprising Two Cellulase
; FILE REFERENCE: 6191 Sequence Listing (8 Sequences)
; Patent No. 6322595
; CURRENT APPLICATION NUMBER: US/09/230, 665
; CURRENT FILING DATE: 1999-01-29
; EARLIER APPLICATION NUMBER: 60/023,125
; EARLIER FILING DATE: 1996-07-30
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 2
; LENGTH: 305
; TYPE: PRP
; ORGANISM: Humicola insolens
US-09-230-665-2
```

Query Match 69.9%; Score 1159; DB 4; Length 305;

Best Local Similarity 67.1%; Pred. No. 7.1e-87;
Matches 206; Conservative 44; Mismatches 47; Indels 10; Gaps 4;

```
QY 1 MRSTPYLRTTLAALPLVASAAGSGOSTRYWDCCPKSCAMPKAAVSOPVYACDANFOR 60
Db 1 MRSSPLLSAVVAALPVLALAA--DGRSTRYWDCCPKSCGMAKAPVNOVPFSCNANFOR 58
QY 61 LSDFNVOGSCN-GGSAYSCADQTPMAVNDNLAYGFAATSIAGSESSWCACAYALFTSG 119
Db 59 ITDFDAKSGCEPGGVASCADQTPMAVNDNDFALGFAATSIAGSNAGWCACAYELFTSG 118
QY 120 PVAGKTMVVOSTSTGGDLGSDNOFDIAMPGGVGIFNGCSSOFGGLPGAQYGGISRPDCD 179
Db 119 PVAGKTMVVOSTSTGGDLGSDNHFDLNPGGVGIFDGCPTQFGGLPGAQYGGISRNEDC 178
QY 180 SFPAPLKPCCOMRPFDMFOVADNPTFTFOOVCPAEIYARSGCKRNDSSFP-VFTPPSGG 238
Db 179 RPPDALKEGCIYRPFDMFKANDNPFSEFROYOCPAEIYARTGCRNDGNFPAVOIPSSST 238
QY 239 NGGTGPTSTAPGSGQTS-----PGGSGCTSQKMAQCGGIGPSGCTTCVSGTTCOKLN 292
Db 239 SSPVNOPTSTSTSTSTSSPPVQPTTPSGCTAERWAQCGGNGMSGCTTCVAGSTCTKIN 298
QY 293 DYYSQCL 299
Db 299 DWYHOCCL 305
```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Comugen Ltd.

OM protein - protein search, using sw model

Run on: February 24, 2003, 11:42:08 ; Search time 13 seconds
(without alignments)
714.606 Million cell updates/sec

Title: US-10-007-521-12

Perfect score: 1659
Sequence: 1 MKSTPVLRTLAALPLVNS.....TTVCSTGTCQKINDYSSQL 299

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 156504 seqs, 31069816 residues

Total number of hits satisfying chosen parameters: 156504

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Maximum Match 0%
Listing first 45 summaries

Database : Published_Applications_AA:*

- 1: /cgn2_6/ptodata/1/pubppaa/US08_NEW_PUB.pep:*
- 2: /cgn2_6/ptodata/1/pubppaa/PCT_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/1/pubppaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/1/pubppaa/US07_NEW_PUB.pep:*
- 5: /cgn2_6/ptodata/1/pubppaa/US07_PUBCOMB.pep:*
- 6: /cgn2_6/ptodata/1/pubppaa/US07_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/1/pubppaa/PCT07_PUBCOMB.pep:*
- 8: /cgn2_6/ptodata/1/pubppaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/1/pubppaa/US09_NEW_PUB.pep:*
- 10: /cgn2_6/ptodata/1/pubppaa/US09_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/1/pubppaa/US10_NEW_PUB.pep:*
- 12: /cgn2_6/ptodata/1/pubppaa/US10_PUBCOMB.pep:*
- 13: /cgn2_6/ptodata/1/pubppaa/US60_NEW_PUB.pep:*
- 14: /cgn2_6/ptodata/1/pubppaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1159	69.9	305	10 US-09-735-787-2	Sequence 2, Appl1
2	1039	62.6	235	1 US-08-841-636A-31	Sequence 31, Appl1
3	939.5	56.6	376	10 US-09-735-787-4	Sequence 4, Appl1
4	160.5	9.7	480	10 US-09-823-936-16	Sequence 16, Appl1
5	160	9.6	33	9 US-09-916-94A-6	Sequence 6, Appl1
6	158	9.5	34	1 US-08-841-636A-37	Sequence 37, Appl1
7	154	9.3	4440	9 US-10-174-590-525	Sequence 525, App
8	154	9.3	4440	9 US-10-176-758-525	Sequence 525, App
9	154	9.3	4440	9 US-10-175-737-525	Sequence 525, App
10	154	9.3	4440	9 US-10-173-706-525	Sequence 525, App
11	154	9.3	4440	9 US-10-175-738-525	Sequence 525, App
12	154	9.3	4440	9 US-10-175-752-525	Sequence 525, App
13	154	9.3	4440	9 US-10-176-482-525	Sequence 525, App
14	154	9.3	4440	9 US-10-176-757-525	Sequence 525, App
15	154	9.3	4440	9 US-10-176-913-525	Sequence 525, App
16	154	9.3	4440	9 US-10-180-552-525	Sequence 525, App
17	154	9.3	4440	9 US-10-180-557-525	Sequence 525, App
18	154	9.3	4440	9 US-10-173-700-525	Sequence 525, App
19	154	9.3	4440	9 US-10-174-572-525	Sequence 525, App

20	154	9.3	4440	9 US-10-174-579-525	Sequence 525, App
21	154	9.3	4440	9 US-10-174-582-525	Sequence 525, App
22	154	9.3	4440	9 US-10-174-588-525	Sequence 525, App
23	154	9.3	4440	9 US-10-175-739-525	Sequence 525, App
24	154	9.3	4440	9 US-10-175-740-525	Sequence 525, App
25	154	9.3	4440	9 US-10-175-743-525	Sequence 525, App
26	154	9.3	4440	9 US-10-176-488-525	Sequence 525, App
27	154	9.3	4440	9 US-10-176-492-525	Sequence 525, App
28	154	9.3	4440	9 US-10-176-747-525	Sequence 525, App
29	154	9.3	4440	9 US-10-176-750-525	Sequence 525, App
30	154	9.3	4440	9 US-10-176-985-525	Sequence 525, App
31	154	9.3	4440	9 US-10-176-987-525	Sequence 525, App
32	154	9.3	4440	9 US-10-176-991-525	Sequence 525, App
33	154	9.3	4440	9 US-10-176-992-525	Sequence 525, App
34	154	9.3	4440	9 US-10-176-993-525	Sequence 525, App
35	154	9.3	4440	9 US-10-184-658-525	Sequence 525, App
36	154	9.3	4440	9 US-10-173-695-525	Sequence 525, App
37	154	9.3	4440	9 US-10-173-697-525	Sequence 525, App
38	154	9.3	4440	9 US-10-173-705-525	Sequence 525, App
39	154	9.3	4440	9 US-10-174-576-525	Sequence 525, App
40	154	9.3	4440	9 US-10-174-585-525	Sequence 525, App
41	154	9.3	4440	9 US-10-174-586-525	Sequence 525, App
42	154	9.3	4440	9 US-10-175-747-525	Sequence 525, App
43	154	9.3	4440	9 US-10-176-481-525	Sequence 525, App
44	154	9.3	4440	9 US-10-176-485-525	Sequence 525, App
45	154	9.3	4440	9 US-10-176-487-525	Sequence 525, App

ALIGNMENTS

RESULT 1
US-09-735-787-2
; Sequence 2, Application US/09735787
; Patent No. US20010036910A1
GENERAL INFORMATION:
APPLICANT: Rasmussen, Grethe
Mikkelsen, Jan Moller
Schulein, Martin
Patzke, Shankant A.
Hagen, Fred
TITLE OF INVENTION: A Cellulase Preparation Comprising an
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. US20010036910A10 NO. US20010036910A1disk of NO. US200100
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/735,787
FILING DATE: 13-Dec-2000
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/189,028
FILING DATE: <unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 3469,214-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 305 amino acids


```

DB 309 --GGGC-----TGC-----CACGAGAGTGTG 329
QY 244 TPTSTAPSGSGTSPGGSG-CTSQKMAOCGGIGFSGCTTVSGTTC 288
DB 330 ACACCTCGAGAAACGGGTTCCAATGCACGTTGC--CTTCACAGTTC 373

```

RESULT 11

```

US-10-175-738-525
; Sequence 525, Application US/10175738
; Publication No. US20030022294A1
; GENERAL INFORMATION:

```

```

; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C45
; CURRENT APPLICATION NUMBER: US/10/175,738
; PRIOR APPLICATION REMOVED - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 525
; LENGTH: 4440
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-175-738-525

```

```

Query Match
Best Local Similarity 24.8%; Score 154; DB 9; Length 4440;
Matches 71; Conservative 19; Mismatches 114; Indels 82; Gaps 14;

```

```

QY 9 TTIAALPLVSAASGSGSTRYWDCKPSCAMPKRAVSQPYACDANFORLSDFNVQS 68
DB 164 TGTGTATGCATGCATGTGTG---CGTGCA-----TGTGTTT 204
QY 69 GCNGSAYSCADQTPMAVNDNLAYGF--AATSIAGSESSMCCACVALFTSGPVAQKTM 126
DB 205 GCGTGTA--CATGTGCATGTGTGTGTGTGCTGTGTGCAC--ATGTAGAAAGAAA 260
QY 127 VVOSTGTGDLGSDNFDIAMPGGGVGIFNGCSSO--FGGLPGAQYGGISSRDQDSFPA 183
DB 261 GTGATGTGTGGGA-----ATGGAAGATTCCAAACCCAGGAGGA--GACTGTGCTG--- 308
QY 184 PLKPGCQWRFDWFMQNDNPTFTFOYQCPAELIYARSGCKRNDSSFVFTPPSGNGGTG 243
DB 309 --GGGC-----TGC-----CACGAGAAAGTGTG 329
QY 244 TPTSTAPSGSGTSPGGSG-CTSQKMAOCGGIGFSGCTTVSGTTC 288
DB 330 ACACCTCGAGAAACGGGTTCCAATGCACGTTGC--CTTCACAGTTC 373

```

RESULT 12

```

US-10-175-752-525
; Sequence 525, Application US/10175752
; Publication No. US20030022295A1
; GENERAL INFORMATION:

```

```

; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.

```

```

; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C60
; CURRENT APPLICATION NUMBER: US/10/175,752
; PRIOR APPLICATION REMOVED - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 525
; LENGTH: 4440
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-175-752-525

```

```

Query Match
Best Local Similarity 24.8%; Score 154; DB 9; Length 4440;
Matches 71; Conservative 19; Mismatches 114; Indels 82; Gaps 14;

```

```

QY 9 TTIAALPLVSAASGSGSTRYWDCKPSCAMPKRAVSQPYACDANFORLSDFNVQS 68
DB 164 TGTGTATGCATGCATGTGTG---CGTGCA-----TGTGTTT 204
QY 69 GCNGSAYSCADQTPMAVNDNLAYGF--AATSIAGSESSMCCACVALFTSGPVAQKTM 126
DB 205 GCGTGTA--CATGTGCATGTGTGTGTGTGCTGTGTGCAC--ATGTAGAAAGAAA 260
QY 127 VVOSTGTGDLGSDNFDIAMPGGGVGIFNGCSSO--FGGLPGAQYGGISSRDQDSFPA 183
DB 261 GTGATGTGTGGGA-----ATGGAAGATTCCAAACCCAGGAGGA--GACTGTGCTG--- 308
QY 184 PLKPGCQWRFDWFMQNDNPTFTFOYQCPAELIYARSGCKRNDSSFVFTPPSGNGGTG 243
DB 309 --GGGC-----TGC-----CACGAGAAAGTGTG 329
QY 244 TPTSTAPSGSGTSPGGSG-CTSQKMAOCGGIGFSGCTTVSGTTC 288
DB 330 ACACCTCGAGAAACGGGTTCCAATGCACGTTGC--CTTCACAGTTC 373

```

RESULT 13

```

US-10-176-482-525
; Sequence 525, Application US/10176482
; Publication No. US20030022296A1
; GENERAL INFORMATION:

```

```

; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C70
; CURRENT APPLICATION NUMBER: US/10/176,482
; PRIOR APPLICATION REMOVED - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 525
; LENGTH: 4440
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-176-482-525

```

```

Query Match
Best Local Similarity 24.8%; Score 154; DB 9; Length 4440;

```

Best Local Similarity 24.8%; Pred. No. 0.001;

Matches 71; Conservative 19; Mismatches 114; Indels 82; Gaps 14;

```

QY 9 TTTLAALPLVASASGSGSTRYWDCKPSCAMPKRAVSOPIVACDANFORLSDFNVOS 68
DB 164 TGTGTATGATGATGATGTGTG-----CGTGCA-----CAGGTG---TGTTT 204
QY 69 GCNGSAYSCADOTPMVAVNDNLAYGF--AATSIAGSESSWCCACVAFPTSGPVAKTM 126
DB 205 GCCTGTA--CATGTGATGTGTGTGTGTGTGCTGTGTGTGCAC--ATGTAGAAAGAAA 260
QY 127 VVOSTSTGDLGNSQDIAMPGGVGIIFNGCSSQ---FGGLPGAQYGGISSRDQDSFPA 183
DB 261 GTGATGTGTGGGA-----ATGAGAAATTCACCAACGAGAGGA--GACTGTGCTG--- 308
QY 184 PLKPGCOMRFDMFQNDNPTFTFOVOCPAEIYARSGCKRNDSSFPVTPPSGNGGTG 243
DB 309 --GGGC-----TG-----CAGGAGAAAGTGTG 329
QY 244 TPTSTAPSGSGQTSPPGSGG-CTSQKWAQCGGIGFSGCTTCVSGTTC 288
DB 330 ACACCTTCGAGAAACGGGTTCACAAATGCACGTTGC--CTTCACAGTTC 373

```

RESULT 14

US-10-176-757-525

```

; Sequence 525, Application US/10176757
; Publication No. US20030022297A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C86
; CURRENT APPLICATION NUMBER: US/10/176,757
; PRIOR APPLICATION REMOVED - See file wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 525
; LENGTH: 4440
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-176-757-525

```

Query Match 9.3%; Score 154; DB 9; Length 4440;

Best Local Similarity 24.8%; Pred. No. 0.001;

Matches 71; Conservative 19; Mismatches 114; Indels 82; Gaps 14;

```

QY 9 TTTLAALPLVASASGSGSTRYWDCKPSCAMPKRAVSOPIVACDANFORLSDFNVOS 68
DB 164 TGTGTATGATGATGATGTGTG-----CGTGCA-----CAGGTG---TGTTT 204
QY 69 GCNGSAYSCADOTPMVAVNDNLAYGF--AATSIAGSESSWCCACVAFPTSGPVAKTM 126
DB 205 GCCTGTA--CATGTGATGTGTGTGTGTGTGCTGTGTGTGCAC--ATGTAGAAAGAAA 260
QY 127 VVOSTSTGDLGNSQDIAMPGGVGIIFNGCSSQ---FGGLPGAQYGGISSRDQDSFPA 183
DB 261 GTGATGTGTGGGA-----ATGAGAAATTCACCAACGAGAGGA--GACTGTGCTG--- 308
QY 184 PLKPGCOMRFDMFQNDNPTFTFOVOCPAEIYARSGCKRNDSSFPVTPPSGNGGTG 243
DB 309 --GGGC-----TGC-----CAGGAGAAAGTGTG 329

```

```

QY 244 TPTSTAPSGSGQTSPPGSGG-CTSQKWAQCGGIGFSGCTTCVSGTTC 288
DB 330 ACACCTTCGAGAAACGGGTTCACAAATGCACGTTGC--CTTCACAGTTC 373

```

RESULT 15

```

US-10-176-913-525
; Sequence 525, Application US/10176913
; Publication No. US20030022298A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C66
; CURRENT APPLICATION NUMBER: US/10/176,913
; PRIOR APPLICATION REMOVED - See file wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 525
; LENGTH: 4440
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-176-913-525

```

Query Match 9.3%; Score 154; DB 9; Length 4440;

Best Local Similarity 24.8%; Pred. No. 0.001;

Matches 71; Conservative 19; Mismatches 114; Indels 82; Gaps 14;

```

QY 9 TTTLAALPLVASASGSGSTRYWDCKPSCAMPKRAVSOPIVACDANFORLSDFNVOS 68
DB 164 TGTGTATGATGATGATGTGTG-----CGTGCA-----CAGGTG---TGTTT 204
QY 69 GCNGSAYSCADOTPMVAVNDNLAYGF--AATSIAGSESSWCCACVAFPTSGPVAKTM 126
DB 205 GCCTGTA--CATGTGATGTGTGTGTGTGTGCTGTGTGTGCAC--ATGTAGAAAGAAA 260
QY 127 VVOSTSTGDLGNSQDIAMPGGVGIIFNGCSSQ---FGGLPGAQYGGISSRDQDSFPA 183
DB 261 GTGATGTGTGGGA-----ATGAGAAATTCACCAACGAGAGGA--GACTGTGCTG--- 308
QY 184 PLKPGCOMRFDMFQNDNPTFTFOVOCPAEIYARSGCKRNDSSFPVTPPSGNGGTG 243
DB 309 --GGGC-----TGC-----CAGGAGAAAGTGTG 329

```

Search completed: February 24, 2003, 11:50:02

Job time : 17 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 24, 2003, 11:49:49 ; Search time 18 seconds
(without alignments)
1596.900 Million cell updates/sec

Title: US-10-007-521-12

Perfect score: 1659

Sequence: 1 MRSTPVLRTTLAALPLVAS.....TTVCSTGTCQKNDYRSQL 299

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR73:*

1: PIR1:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	740.5	44.6	229	2 JC7308	cellulase (EC 3.2.1.4)
2	507	30.6	511	2 SI0527	endoglucanase B pr
3	462	27.8	393	2 S59499	cellulase eg11
4	221	13.3	242	2 S60143	cellulase (EC 3.2.1.4)
5	219.5	13.2	513	1 EUP01	cellulase 1,4-beta
6	219.5	13.2	513	2 S45360	cellulase 1,4-beta
7	215.5	13.0	513	2 S11439	cellulase 1,4-beta
8	214	12.9	459	2 A25928	cellulase (EC 3.2.1.4)
9	203	12.2	463	2 A48375	cellulase (EC 3.2.1.4)
10	190	11.5	516	2 S33164	cellulase 1,4-beta
11	189	11.4	464	2 JC7143	cellulase 1,4-beta
12	188.5	11.4	302	2 S71334	endoglucanase I -
13	185	11.2	525	2 S38794	acetyl xylan ester
14	184	11.1	516	2 JS0083	cellulase 1,4-beta
15	183	11.0	537	2 JU0150	cellulase 1,4-beta
16	181	10.9	516	2 S42093	cellulase 1,4-beta
17	172	10.4	320	2 JC1311	cellulase 1,4-beta
18	170.5	10.3	510	2 S41943	cellulase 1,4-beta
19	170.5	10.3	511	2 S44716	cellulase 1,4-beta
20	161	9.7	504	2 B48939	cellulase 1,4-beta
21	156	9.4	856	2 T00349	cellulase III - As
22	150.5	9.1	860	2 S43846	cellulase B - rumen
23	143.5	8.6	388	2 JC5461	cellulase (EC 3.2.1.4)
24	143.5	8.6	388	2 S43920	cellulase (EC 3.2.1.4)
25	139.5	8.4	540	2 S41942	cellulase 1,4-beta
26	137	8.3	438	1 S70602	cellulase 1,4-beta
27	137	8.3	839	2 S75518	hypothetical prote
28	129	7.8	418	1 S28372	cellulase (EC 3.2.1.4)
29	127	7.7	429	1 JC5861	endo-1,4-beta-xyla

30	125.5	7.6	316	1 A38743	loriferin - human
31	125.5	7.6	410	1 S68153	cellulase (EC 3.2.1.4)
32	124	7.5	471	1 A26160	cellulase 1,4-beta
33	124	7.5	471	1 A38979	cellulase 1,4-beta
34	122	7.4	182	2 A36686	ultra-high-sulfur
35	120.5	7.3	303	2 H70716	hypothetical prote
36	120	7.2	5376	2 T42215	zonadhesin - mouse
37	118.5	7.1	453	2 A41640	vestigial protein
38	118.5	7.1	749	2 I38488	trophinin - human
39	116.5	7.0	584	2 G70804	hypothetical glycl
40	115.5	7.0	191	2 I46412	keratin KAP5.4 - s
41	114	6.9	461	2 T51044	related to spore c
42	114	6.9	1804	2 H96597	hypothetical glycl
43	113.5	6.8	1381	2 E70806	hypothetical glycl
44	113.5	6.8	2090	2 S26058	probable transform
45	111.5	6.7	1032	2 T34433	hypothetical prote

ALIGNMENTS

RESULT 1
JC7308
cellulase (EC 3.2.1.4) - Scopulariopsis brevicaulis
N:Alternate names: endoglucanase I
C:Species: Scopulariopsis brevicaulis
C:Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 18-Aug-2000
C:Accession: JC7308; PC7087
R:Nakatani, F.; Kawaguchi, T.; Takada, G.; Sumitani, J.; Moriyama, Y.; Arai, M.
B:Sci. Biotechnol. Biochem. 64, 1238-1246, 2000
A:Title: Cloning and sequencing of an endoglucanase gene from Scopulariopsis brevicaulis
A:Reference number: JC7308
A:Accession: JC7308
A:Molecule type: DNA
A:Residues: 1-229 <NA>
A:Experimental source: strain TOF-1212
A:Accession: PC7087
A:Molecule type: protein
A:Residues: 21-37;149-164 <NA>
C:Genetics:
A:Gene: eg1
A:Introns: 147/3
C:Keywords: glycosidase; hydrolase

Query Match 44.6%; Score 740.5; DB 2; Length 229;
Best local similarity 59.8%; Pred. No. 2.6e-45;
Matches 134; Conservative 32; Mismatches 49; Indels 9; Gaps 5;

OY 10 TLAALPLV---ASAAGSGGSTRYWDCKPSCAMPKRAAVSO-PVYACDANFORLSPN 65
DB 6 TLLALPLVLPRAASQSGTGTTTRIMWDCCKPSCWPDRAPLSQGPMTCDINDNPLDDG 65
OY 66 V-OQSGCN-GGSAYSCADQTPMAVNDNLAYGFAATSIAGSESSSMCCACALFTTSGPVAG 123
DB 66 LTESCEPEGGAGYMCSSHPAVDDELAYGMAAVNIIGGTESDMCCACALEFTTGAWSG 125
OY 124 KTMVQSTSTGSDGSGNPDIAIPGCGVGFNCGSSQFGLP---GAOTGIGSSDDQDS 180
DB 126 KMLVQATNTGSDGNNHFDIAMPGGVGFNCGSDQSPNGSGDRGVHTRADDS 185
OY 181 FPAFLKPGCOWRFQWADNPTFFQOVQCPAETIARSGCKRN 224
DB 186 FPEALKACGEMRFDFWFGTDPNDVVSFRVECPALVQKSCORS 229

RESULT 2
SI0527
endoglucanase B precursor - Pseudomonas fluorescens
C:Species: Pseudomonas fluorescens
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 08-Oct-1999
R:Gilbert, H.J.; Hall, J.; Hazlewood, G.P.; Ferreira, L.M.A.
Mol. Microbiol. 4, 759-767, 1990

RESULT 5

EURL1

cellulose 1,4-beta-cellulobiosidase (EC 3.2.1.91) I precursor - fungus (Trichoderma reesei)

C:Species: Trichoderma reesei

C>Date: 03-Aug-1984 #sequence_revision 03-Aug-1984 #text_change 16-Jul-1999

C:Accession: A00902

R:Shomaker, S.; Schweickart, V.; Ladner, M.; Gelfand, D.; Kwok, S.; Myambo, K.; Innis, B./Technology 1, 691-696, 1993

A:Title: Molecular cloning of exo-cellulohydrolase I derived from Trichoderma reesei

A:Reference number: A00902

A:Accession: A00902

A:Molecule type: DNA

A:Residues: 1-513 <SHO>

A:Experimental source: strain L27

C:Comment: This is the most abundantly produced cellulase in this filamentous fungus; it

C:Genetics:

A:Gene: CBH1

A:Introns: 154/2; 386/3

C:Function:

A:Description: catalyzes the hydrolysis of 1,4-beta-D-glucosidic bonds in cellulose to D

C:Superfamily: cellulose 1,4-beta-cellulobiosidase I; fungal cellulose-binding domain hom

C:Keywords: glycosidase; hydrolase; polysaccharide degradation

F:1-17/Domain: signal sequence #status predicted <SIG>

F:18-513/Product: cellulose 1,4-beta-cellulobiosidase I #status predicted <MAT>

F:482-513/Domain: fungal cellulose-binding domain homology <FCB>

Query Match

Best local Similarity 13.2%; Score 219.5; DB 1; Length 513;
Matches 91; Conservative 36; Mismatches 122; Indels 89; Gaps 22;

16 PLVSAASGSGSTRYWDCKPSCAMPKGAASQPV--VACDANFQRLSDFNVOGCGNG 73

211 PSSNNANTGIG---GHGSCSEMDIWEAN-SISEALPHRCTTVGQICE---GDGC--G 261

74 SAYS-----C-----ADQTPMAVNDNLAYGFAATSIAGSESSMCCACALFTTSPVA 122

262 GYSDNRNGGTCDDPGCDMNPYRLGNTSFYG-----PGSS-----FTLDITTK----- 303

123 GKRWVOSTSTGDLG---SNQFDIAMPGGGVIFNG-----CS---SQFGCLPGAQ 168

304 -KLTVVTFQFETSGAINRYVQNGVTFQGPNAELGSGNELNDYCTAEAEAEFGSSFSFSD 362

169 YGSISSRDCCDFFAPLPGCCQMFDFQON---ADNPTFTFQOVQCPAEIYARSCKRND 225

363 KGGITQPKKATSGGMYLVMSL-WD-DYANMLMLDSTYPTNETSSTPGAV--RGSCTSS 418

226 -----DSFPP-----VFPP-----PSGNN---GGTGTPTSTAPSGQTSPPGGS 261

419 GVPAYVESQSPNAKVFYSNFKFPIGSTGNPSGNGPPGCGRGTTRRPA---TTTGSSP 475

262 GCTSQKMAQCGIGFSGCTTCVSGTTCCQKINDYSSQL 299

476 GPTQSHYGCGGIGSGPTVCASGTTQVLPNPPYSQCL 513

RESULT 6

cellulose 1,4-beta-cellulobiosidase (EC 3.2.1.91) - fungus (Trichoderma koningii)

C:Species: Trichoderma koningii

C>Date: 20-Oct-1994 #sequence_revision 10-Nov-1995 #text_change 22-Jun-1999

C:Accession: S45380

R:Wey, T.T.; Heeu, T.H.; Huang, L.

C:Microbiol. 28, 31-39, 1994

A:Title: Molecular cloning and sequence analysis of the cellulohydrolase I gene from Tr

A:Reference number: S45380; MUID:94100788; PMID:7764306

A:Accession: S45380

A:Molecule type: DNA

A:Residues: 1-513 <KEY>

A:Cross-references: EMBL:X69976; NID:9457422; PIDN:CAA9596.1; PID:9457423

C:Genetics:

A:Introns: 154/2; 386/3
C:Superfamily: cellulose 1,4-beta-cellulobiosidase I; fungal cellulose-binding domain h
C:Keywords: glycosidase; hydrolase; polysaccharide degradation
F:482-513/Domain: fungal cellulose-binding domain homology <FCB>

Query Match

Best local Similarity 13.2%; Score 219.5; DB 2; Length 513;
Matches 91; Conservative 36; Mismatches 122; Indels 89; Gaps 22;

16 PLVSAASGSGSTRYWDCKPSCAMPKGAASQPV--VACDANFQRLSDFNVOGCGNG 73

211 PSSNNANTGIG---GHGSCSEMDIWEAN-SISEALPHRCTTVGQICE---GDGC--G 261

74 SAYS-----C-----ADQTPMAVNDNLAYGFAATSIAGSESSMCCACALFTTSPVA 122

262 GYSDNRNGGTCDDPGCDMNPYRLGNTSFYG-----PGSS-----FTLDITTK----- 303

123 GKRWVOSTSTGDLG---SNQFDIAMPGGGVIFNG-----CS---SQFGCLPGAQ 168

304 -KLTVVTFQFETSGAINRYVQNGVTFQGPNAELGSGNELNDYCTAEAEAEFGSSFSFSD 362

169 YGSISSRDCCDFFAPLPGCCQMFDFQON---ADNPTFTFQOVQCPAEIYARSCKRND 225

363 KGGITQPKKATSGGMYLVMSL-WD-DYANMLMLDSTYPTNETSSTPGAV--RGSCTSS 418

226 -----DSFPP-----VFPP-----PSGNN---GGTGTPTSTAPSGQTSPPGGS 261

419 GVPAYVESQSPNAKVFYSNFKFPIGSTGNPSGNGPPGCGRGTTRRPA---TTTGSSP 475

262 GCTSQKMAQCGIGFSGCTTCVSGTTCCQKINDYSSQL 299

476 GPTQSHYGCGGIGSGPTVCASGTTQVLPNPPYSQCL 513

RESULT 7

S11439

cellulose 1,4-beta-cellulobiosidase (EC 3.2.1.91) - fungus (Trichoderma viride)

C:Species: Trichoderma viride

C>Date: 21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 22-Jun-1999

C:Accession: S11439

R:Cheng, C.; Tsukagoshi, N.; Ueda, S.

Nucleic Acids Res. 18, 5559, 1990

A:Title: Nucleotide sequence of the cellulohydrolase gene from Trichoderma viride.

A:Reference number: S11439; MUID:91016856; PMID:2216737

A:Accession: S11439

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-513 <CHE>

A:Cross-references: EMBL:X53931; NID:95196; PIDN:CAA37878.1; PID:9295937

C:Genetics:

A:Introns: 154/2; 386/3

C:Superfamily: cellulose 1,4-beta-cellulobiosidase I; fungal cellulose-binding domain h

C:Keywords: glycosidase; hydrolase; polysaccharide degradation

F:482-513/Domain: fungal cellulose-binding domain homology <FCB>

Query Match

Best local Similarity 13.0%; Score 215.5; DB 2; Length 513;
Matches 86; Conservative 36; Mismatches 137; Indels 69; Gaps 16;

16 PLVSAASGSGSTRYWDCKPSCAMPKGAASQPV--VACDANFQRLSDFNVOGCGNG 73

211 PSSNNANTGIG---GHGSCSEMDIWEAN-SISEALPHRCTTVGQICEGDSGGTYSG 266

74 SAY--SC-----ADQTPMAVNDNLAYGFAATSIAGSESSMCCACALFTTSGPAGKTMV 127

267 DRFGGCTDDPGCDMNPYRLGNTSFYG-----PGSS-----FTLDITTK-----RLTV 307

128 VQSTSTGDLG---SNQFDIAMPGGGVIFNGS-----SQFGCLPGAQYGIS 173

308 VTQFETSGAINRYVQNGVTFQGPNAELGSGSLDDYCAAEAEAEFGSSFSFSDKGLT 367

174 SRDCCSFPAPLPGCCQMFDFQON---ADNPTFTFQOVQCPAEI---VARSSCKRND 226

Db 368 OFKATSGMVLVMSL-WD-DYANMLMDSTYPTDETSTPGAVRGSSSTSSGVAOLE 425

QY 227 SSFP-----VFPP-----PSGNGGTCTPTSTAPSGGQTSPPGGSGCTSGKMAOC 271

Db 426 SNSPMNAKVYSNIRKFGIGSTGNPSGPNPGNPPGTTTPPAPVSTGSSPGPTQHHYGC 485

QY 272 GGIGFSGCTTCVSGTTCOKLNDYYSOCL 299

Db 486 GGIGYIGPYVCASSTGCVLNPYISOCL 513

RESULT 8
A25928
cellulase (EC 3.2.1.4) A precursor - fungus (*Trichoderma reesei*)

N:Alternate names: endo-1,4-beta-glucanase; endoglucanase I
C:Species: *Trichoderma reesei*
C>Date: 15-Dec-1988 #sequence_revision 15-Dec-1988 #text_change 22-Jun-1999
C:Accession: A25928; A25565
R:Van Arsdel, J.N.; Kwok, S.; Schweickart, V.L.; Ladner, M.B.; Gelfand, D.H.; Innis, M.
Bio/Technology 5, 60-64, 1987
A:Title: Cloning, characterization, and expression in *Saccharomyces cerevisiae* of endoglucanase
A:Reference number: A25928
A:Accession: A25928
A:Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-459 <VAN>
A:Experimental source: strain L27
R:Penlidze, M.; Lehtovaara, P.; Nevalainen, H.; Bhikhabhai, R.; Knowles, J.
Gene 45, 253-263, 1986
A:Title: Homology between cellulase genes of *Trichoderma reesei*: complete nucleotide sequence
A:Reference number: A25565; MUID:87106822; PMID:2948877
A:Accession: A25565
A:Molecule type: mRNA
A:Residues: 1-459 <PEN>
A:Cross-references: GB:M15665; NID:g170546; PID:AAA34212.1; PID:g170547
C:Genetics:
A:Gene: eg11
A:Introns: 257/2; 454/1
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cel-
lulose
A:Pathway: cellulose degradation
C:Superfamily: cellulase 1,4-beta-cellulobiosidase I; fungal cellulose-binding domain hom-
C:Keywords: glycosidase; hydrolase; polysaccharide degradation
F:1-22/Domain: signal sequence #status predicted <SIG>
F:23-459/Product: cellulase A #status predicted <MAT>
F:428-459/Domain: fungal cellulose-binding domain homology <FCB>

Query Match 12.9%; Score 214; DB 2; Length 459;
Best local similarity 27.0%; Pred. No. 6e-08;
Matches 108; Conservative 31; Mismatches 101; Indels 160; Gaps 27;

QY 19 ASASGSGS---TRYMDCCKPCAMPKAAVSOPVYACDA-----NFQRISDENP 66

Db 101 ASGVYSGSSLTNNQY---WPSSS-GGYSSVSPRLYLDSGCEYVMLKLNGQELS-FPV 154

QY 67 Q-----SGCNGSAYSCADQTP---WAVNDNLAYGFA 95

Db 155 DLSALPGGENSLTSLQMDENGANQYTAGANTGSGY-CDAGCPVQYWR-NGTL----- 207

QY 96 ATSTAGSESSWC-----AC-----YALFTTS--GPV 121

Db 208 -----NTSHQFCNEMDILEGNSRANALTPHSCTATACDSAGCFNPGSGYKSYGPP- 261

QY 122 AGKTMVQST-----STGCDLGS-----NQFDI--AMPGGVGIIFNGCSSQ 160

Db 262 -GDVYDTSKPTTTITQFNTDNGSPSGNLVSTIRKYQNGVDIPSAQPGDT--ISSCP- 317

QY 161 FGLPGAGYGGISSRDQDSFPAPLPCQWRFD-WFONAD--NPTFTTQOYQC----- 211

Db 318 -----ASATYGLATMGKA-----LSSGMVLVFSITMNDNSQYMNMLDGSNAPPCSTEEN 366

QY 212 PAETIVASGCKRNDSSFPVFTPPSGNGGCTGTPTSTA---PGSGQT-----SPGG 259

Db 367 PSNIIA-----NNPNTHVVESNIRMGDGS-TTNSTAPPPEPPASSTFTTTRSSSTTS 419

QY 260 GSGCTSGKMAOCGGIGFSGCTTCVSGTTCOKLNDYYSOCL 299

Db 420 SPSCOTIHWGCGGIGISGCKTCTSGTTCOYSNDYYSOCL 459

RESULT 9
A48375
cellulase (EC 3.2.1.4) - fungus (*Trichoderma longibrachiatum*)

N:Alternate names: endo-1,4-beta-glucanase; endoglucanase I homolog Eg11
C:Species: *Trichoderma longibrachiatum*
C>Date: 19-Nov-1993 #sequence_revision 18-Nov-1994 #text_change 22-Jun-1999
C:Accession: A48375; S28521
R:Gonzalez, R.; Ramon, D.; Perez-Gonzalez, J.A.
Appl. Microbiol. Biotechnol. 38, 370-375, 1992
A:Title: Cloning, sequence analysis and yeast expression of the eg11 gene from *Tricho-*
A:Reference number: A48375; MUID:93159747; PMID:1369161
A:Accession: A48375
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-463 <GON>
A:Cross-references: EMBL:X60652; NID:g5181; PID:CAA43059.1; PID:g5182
A:Note: sequence extracted from NCBI backbone (NCBIN:125157, NCBIPI:125158)
C:Genetics:
A:Introns: 257/2; 458/1
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as
cellulose
A:Pathway: cellulose degradation
C:Superfamily: cellulase 1,4-beta-cellulobiosidase I; fungal cellulose-binding domain h
C:Keywords: extracellular protein; glycosidase; hydrolase; polysaccharide degradation
F:432-463/Domain: fungal cellulose-binding domain homology <FCB>

Query Match 12.2%; Score 203; DB 2; Length 463;
Best local similarity 28.4%; Pred. No. 3.6e-07;
Matches 86; Conservative 22; Mismatches 101; Indels 94; Gaps 18;

QY 68 SGCNGSAYSCADQTP---WAVNDNL---AYGFAAT---SIAGSESSWCACVATFTTSG 119

Db 184 AGANYSGSY-CDAGCPVQYWR-NGTLMTSGGFCFCNEMDILEGNSRANALPHSOTATAC 241

QY 120 PVAG-----KTMVY-----QSTSTGDLGS-----NQFD 143

Db 242 DSAGCGFNPGSGYPNFSGGDTVDTSKFTIITQFNTDNGSPSGNLVSTIRKYRONQVD 301

QY 144 I--AMPGGVGIIFNGCSSQPGGLPGAQYGGISSRDQDSFPAPLPCQWRFD-WFONAD 200

Db 302 IPSAKPQGDY--ISSCPS-----ASATYGLATMGKA-----LSSGMVLVFSITMNDNSQ 347

QY 201 -----NPTFTTQOYQCPAEIVARSCKRNDSSFPVFTPPSGN--- 239

Db 348 YNMWLDSSGRAPCSSTEGNSNII--ANNPETHVYVSIRMGDLS---TNTSTGNNPP 402

QY 240 ---GGTGTPTSTAPSGQTSPPGGSGCTSGKMAOCGGIGFSGCTTCVSGTTCOKLNDYY 296

Db 403 PPPASSTFTSTTRRSSTTS--SSPSCOTIHWGCGGIGYGCXTCXSGTTCQYNDYYS 460

QY 297 QCL 299

Db 461 QCL 463

RESULT 10
S33164
cellulose 1,4-beta-cellulobiosidase (EC 3.2.1.91) - basidiomycete (*Phanerochaete chrysos-*
C:Species: *Phanerochaete chrysosporium*
C>Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 22-Jun-1999
C:Accession: S33164
R:Slims, P.F.G.; Soares-Felipe, S.M.; Gent, M.E.; Tempelaars, C.; Wang, Q.; Broda, P.
submitted to the EMBL data library, April 1993
A:Description: Differential expression of multiple exo-cellulohydrolase I-like genes

A:Accession: S33164
A:Molecule type: DNA
A:Residues: 1-516 <STM>
A:Cross-references: EMBL:J22528; NID:g296026; PIDN:CA80253.1; PID:g296027
C:Superfamily: cellulose 1,4-beta-cellobiosidase I; fungal cellulose-binding domain homology
C:Keywords: glycosidase; hydrolase; polysaccharide degradation
F:405-516/domain: fungal cellulose-binding domain homology <FCB>

Query Match	11.5%;	Score 190;	DB 2;	Length 516;
Best Local Similarity	24.4%;	Pred. No. 3.3e-06;		
Matches	88;	Conservative	34;	Mismatches 101;
				Indels 138;
				Gaps 18;

QY	20	SAASGSGSTRYYMDCCKPSCAMPK--AAVSOPVACDANFRQLSDPNVSGCGG----	73
		: :	: :
Db	211	SANAGTC--NYGCGCTEMDIWEANNDAAATP-HPCIT-----NAOTRCGSQCTR	256
		: :	: :
QY	74	-----SAYSCADOT-----PMAY-----NDNLAVGFATSIAGS	103
		::: :	::: :
Db	259	DTGLCADGCDENFSFRMGDDFTLKGTLVDTSKREFTVQTGTIDGTSAG-----	308
		: :	: :
QY	104	ESSMCCACIALTFISGVACGKTMVOSTS-----GGDLGSMQFDIAMPGGVGIFNGCSSQ	160
		: :	: :
Db	309	-----TLTEIRRLVYONGKVIQNSVXKIPGIDLVNSTITD-----NFCSSQ	348
		: :	: :
QY	161	-----FGGLP-GAOVGGISRDCC-----DSPEAPLKPCCQMRPFMQ-----NA	199
		: :	: :
Db	349	KTAGAGDINTFQAQHGGLKVOGBALRTGAVLALSTWDYDIAANML-----WLDSSNPYTK	400
		: :	: :
QY	200	DNPTFEQOVOC-----PAIVARSCKRNDSSPFVFPSPSGNGCT-----	242
		: :	: :
Db	401	DPSTPGVARGKCATTSQVPAQIENS-----PAAVYVESNINKEGDLNTTYTGIVSSSV	454
		: :	: :
QY	243	-----GPTSTAPESGCTSPGSGSCTSQKWAQCGIGSGCTTCVSGTTQCKRLNDYSSQ	297
		: :	: :
Db	455	SSSHSSTSSSSSSSTPTPOPTGVYVPMQCGCGGIGYGTSCASPYTCHVLNPPYSQ	514
		: :	: :
QY	298	-----C	298
		: :	: :
Db	515	C	515

RESULT 11
JC7143
endoglucanase I - fungus (*Trichoderma viride*)
C:Species: *Trichoderma viride*
C:Date: 04-Mar-2000 #sequence_revision 04-Mar-2000 #text_change 09-Jun-2000
C:Accession: JC7143
R:Kwon, I.; Ekino, K.; Goto, M.; Furukawa, K.
Biosci. Biotechnol. Biochem. 63, 1714-1720, 1999
A:Title: Heterologous expression and characterization of endoglucanase I (Egi) from *Trichoderma reesei*
A:Reference number: JC7143; MUID:20052947; PMID:10586500
A:Accession: JC7143
A:Molecule type: mRNA
A:Residues: 1-464 <KMO>
A:Experimental source: HK-75
C:Comment: This protein is a fusion glycoprotein with catalytic and cellulose binding domains.
C:Genetics:
A:Gene: egli
A:Introns: 257/2; 459/1
C:Superfamily: cellulose 1,4-beta-cellulobiosidase I; fungal cellulose-binding domain homolog
C:Keywords: fusion protein; glycoprotein
C:453-464/Domain: fungal cellulose-binding domain homology <FCBA>

Query Match	11.4%;	Score 189;	DB 2;	Length 464;
Best Local Similarity	25.4%;	Pred. NO. 3.5e-06;		
Matches 102;	Conservative 32;	Mismatches 108;	Indels 160;	Gaps 25;

```

0Y      21 AASG---SGGS-----TRWDCCKPSCAMPGRANAASQPYIACDA-----NFORLSDFN 65
          ||||| | | : | | | | | : | | | | | : | | | | | :
Db     100 AAGCVOTSGASLTMNOY---MPSSS-GGVSVSPRLYYLLDSDGEVYLKLNCGELS-FD 153
          ||||| | | : | | | | | : | | | | | : | | | | | :
0Y      66 VQ-----SGCNGSAYSCADDTP---WA>VNDNL---A 91
          | | | | | | | | | | | | | | | | | | | | | |

```

[illegible]

RESULT 12
S71334
acetyl xylan esterase precursor - fungus (*Trichoderma reesei*)
C:Species: *Trichoderma reesei*
C>Date: 23-Jul-1997 #sequence_revision 01-Aug-1997 #tex_change 17-Mar-1999
C/Accession: S71334
R:Margolles-Clair, E.; Tenkanen, M.; Soederlund, H.; Penttiläe, M.
Eur. J. Biochem. 237, 553-560, 1996
A>Title: Acetyl xylan esterase from *Trichoderma reesei* contains an active-site serine
A/Reference number: S71334; MUID:96255218; PMID:8647098
A/Accession: S71334
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-302 <MAR>
A/Cross-references: EMBL:Z69256; NID:g1431619; PID:e220701; PID:g1431620
C/Genetics:
A:Gene: axel
C:Superfamily: fungal cellulose-binding domain homology
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-302/Product: acetyl xylan esterase #status predicted <MAT>
F:271-302/Domain: fungal cellulose-binding domain homology <FCB>

Query Match	11.4%	Score 188.5	DB 2	Length 302
Best Local Similarity	26.6%	Pred. No. 2.5e-06		
Matches 73	Conservative 17	Mismatches 97	Indels 87	Gaps 8

[illegible]

RESULT 13
S38794
cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91) - imperfect fungus (*Humicola grisea*)

N:Alternate names: beta-glucanocellulohydrolase; exoglucanase

C:Species: Humicola grisea var. thermoides

C>Date: 10-Sep-1999 #sequence #revision 10-Sep-1999 #text_change 10-Sep-1999

C:Accession: S38794; S08240; A45869

R:Radford, A.

submitted to the EMBL Data Library, June 1991

A:Reference number: S38794

A:Accession: S38794

A:Molecule type: DNA

A:Residues: 1-525 <RAD>

A:Cross-references: EMBL:X17258; NID:g2760; PIDN:CAA3159.1; PID:g2761

A:Note: this is a revision to the sequence from reference S08240

R:de Oliveira Azevedo, M.; Radford, A.

Nucleic Acids Res. 18, 668, 1990

A:Title: Sequence of cbh-1 gene of Humicola grisea var. thermoides.

A:Reference number: S08240; MUID:90175006; PMID:2308855

A:Accession: S08240

A:Molecule type: DNA

A:Residues: 1-299, 'H', 301-525 <DEO>

A:Cross-references: EMBL:X17258

A:Note: the authors translated the codon CAG for residue 87 as His

A:Note: this sequence has been revised in reference S38794

R:Azevedo, M.; de, O.; Felipe, M.S.; Astolfi-Filho, S.; Radford, A.

J. Gen. Microbiol. 136, 2569-2576, 1990

A:Title: Cloning, sequencing and homologues of the cbh-1 (exoglucanase) gene of Humicola

A:Reference number: A45869; MUID:91178527; PMID:2127803

A:Accession: A45869

A:Status: not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 1-20, 'R', 22-34, 'K', 36-86, 'H', 88-141, 'V', 143-157, 'Y', 159-237, 'QOH', 241-244, 'I'

A:Cross-references: GB:M64588; GB:X17258

A:Note: this sequence has been revised. See entry S08240

C:Genetics:

A:Gene: cbh-1

A:introns: 138/1

C:Superfamily: cellulose 1,4-beta-cellulobiosidase I; fungal cellulose-binding domain hom

C:Keywords: glycosidase; hydrolase; polysaccharide degradation

F:494-525/Domain: fungal cellulose-binding domain homology <FCB>

Query Match 11.2%; Score 185; DB 1; Length 525;

Best Local Similarity 24.9%; Pred. No. 7.5e-06;

Matches 97; Conservative 33; Mismatches 136; Indels 124; Gaps 23;

11 LAALPLVASAASGSGSTRYDCCCKPCAMPG-KAAYSQPVYACDANFORLSDF----- 64

159 LAGALYVSMADDSG--LSRY-----PGMKAGAKYGTGYCDACCPRIKFINGEA 206

65 NVSGCN-----GGSAYS--CADQTPMAVNDNLAYGFA--AISIAGS--ESSWCAC 111

207 NIEGWTGSTNDPNAGAGRYGTCCSEMDIWEAN-NMATAFTHPTIIGOSRCEGDSGCGT 265

112 YALFTTSG-----PVAGKTMVYOST-----STGDLGS-NOF 142

266 YSNERYAGVCDPDGCDPNRYGKKTFTGKMTYDTTKITVYVLOFLDANDGLDEIKRF 325

143 -----DIAMPG--GGVGIFNGCSSQ--FGGLPG-AQYGGISSRDQCDSPAPVL 185

326 YVODGKITPNSSESTIPGEGNSITQDMCDRQKVAIFDIDDEFNRKGMOMKALAGPMVL 385

186 KPGCGMRFDFONADNPFF--TFQOVCCPAEIVARSGCKRN--DSSFPV----- 231

386 VMSI-----WDDHASNMLMLDSTF-----PYDAAGKPGAERACPTTSGVPAVEAEADN 435

232 -----FPP-----PSGGNGGTG-----TPSTAPSGQTSPPGSGGCTSOAKWA 269

436 SNVFSNIRECPISITVAGLFGAGNGGNNGNPPPTTTTSSAPATTTTASAGPRAGRWQ 495

270 QCGGIGSGCTTCVSGTQKRLNDYSSQL 299

496 QCGGIGTGTPTQCEEPYICTRLNDYSSQL 525

RESULT 14

JS0083

cellulose 1,4-beta-cellulobiosidase (EC 3.2.1.91) I - basidiomycete (Phanerochaete chry

C:Species: Phanerochaete chrysosporium

C>Date: 31-Mar-1992 #sequence #revision 31-Mar-1992 #text_change 04-Sep-1998

C:Accession: JS0083

R:Stms, P.; James, C.; Broda, P.

Gene 74, 411-422, 1988

A:Title: The identification, molecular cloning and characterisation of a gene from Ph

A:Reference number: JS0083; MUID:89232732; PMID:3246351

A:Accession: JS0083

A:Molecule type: DNA

A:Residues: 1-516 <STM>

A:Experimental source: strain ME446

C:Comment: This protein is a component of the inducible cellulase complex.

C:Genetics:

A:introns: 201/3; 511/1

C:Superfamily: cellulose 1,4-beta-cellulobiosidase I; fungal cellulose-binding domain h

C:Keywords: glycosidase; hydrolase; polysaccharide degradation

F:485-516/Domain: fungal cellulose-binding domain homology <FCB>

F:81.92/Active site: Asp #status predicted

Query Match 11.1%; Score 184; DB 2; Length 516;

Best Local Similarity 24.7%; Pred. No. 8.7e-06;

Matches 86; Conservative 37; Mismatches 113; Indels 112; Gaps 20;

20 SASGSGSTRYDCCCKPCAMPK--AAVSQPVYACDANFORLSDFNVSGCNGG---- 73

211 SANAGTG---NYGTCTCEMDIWEANNDAAATP-HPTT-----NAQTRSGSDCTR 258

74 -----SAYCAQDTPMAVNDNLAYGFAAISIAGSSESSWCACALFTSGPV 121

259 DTGLCADGCDNFSPFMDPT-----FLGKGLITVDTSKFTFVYVQFTI-NGTS 306

122 AG-----KTMVYSTSTGDLGSDNOFDIAMPG--GGVGIF--NGCSSQ--FGGLP-GAAY 169

307 AGITFTIRRLVYGN-----GKVIQNS--SVKIPGIDPVNSITDNCSQKTAFTDNYFAOH 361

170 GGTSRSDQC-----DSFPAPLRKGCQMRFPWQ-----NADNPFTFTQVQC- 211

362 GGLKQGEALRTGMVLALSIMDDYANML-----WLDNYPFTNKPSTPGVARGTCA 413

212 -----PAEIVARSGCKRNDSSFPVFTPPSGNGGT-----GTPSTAP 250

414 TTSGVPAQLIAQS-----PNATVFSNIRECPISITVAGLFGAGNGGNNGNPPPTTTTSSAPATTTTASAGPRAGRWQ 467

251 GSGQTSPPGSGGCTSOAKWACGGIGSGGCTTCVSGTQKRLNDYSSQL 298

468 SSSSTPTPTQGTVPQWCGCGGIGYGTGCAAPYCHVILNPYSQC 515

RESULT 15

cellulose 1,4-beta-cellulobiosidase (EC 3.2.1.91) I - Penicillium janthinellum

N:Alternate names: exo-cellulohydrolase

C:Species: Penicillium janthinellum

C>Date: 10-Sep-1999 #sequence #revision 10-Sep-1999 #text_change 16-Jun-2000

C:Accession: J00150; S21508

R:Christoph, A.K.; Weigel, T.O.; Scholz, G.

Gene 124, 57-65, 1993

A:Title: Cloning, sequencing, and heterologous expression of a cellulase-encoding cDN

A:Reference number: J00150; MUID:93178976; PMID:8440481

A:Accession: J00150

A:Molecule type: mRNA

A:Residues: 1-537 <CHR>

A:Cross-references: EMBL:X59054; NID:g3177; PIDN:CAA41780.1; PID:g3178

A:Note: the authors translated the codon ATG for residue 172 as Asn

C:Genetics:

A:Gene: cbh1

C:Superfamily: cellulose 1,4-beta-cellulobiosidase I; fungal cellulose-binding domain h

C:Keywords: glycosidase; hydrolase; polysaccharide degradation

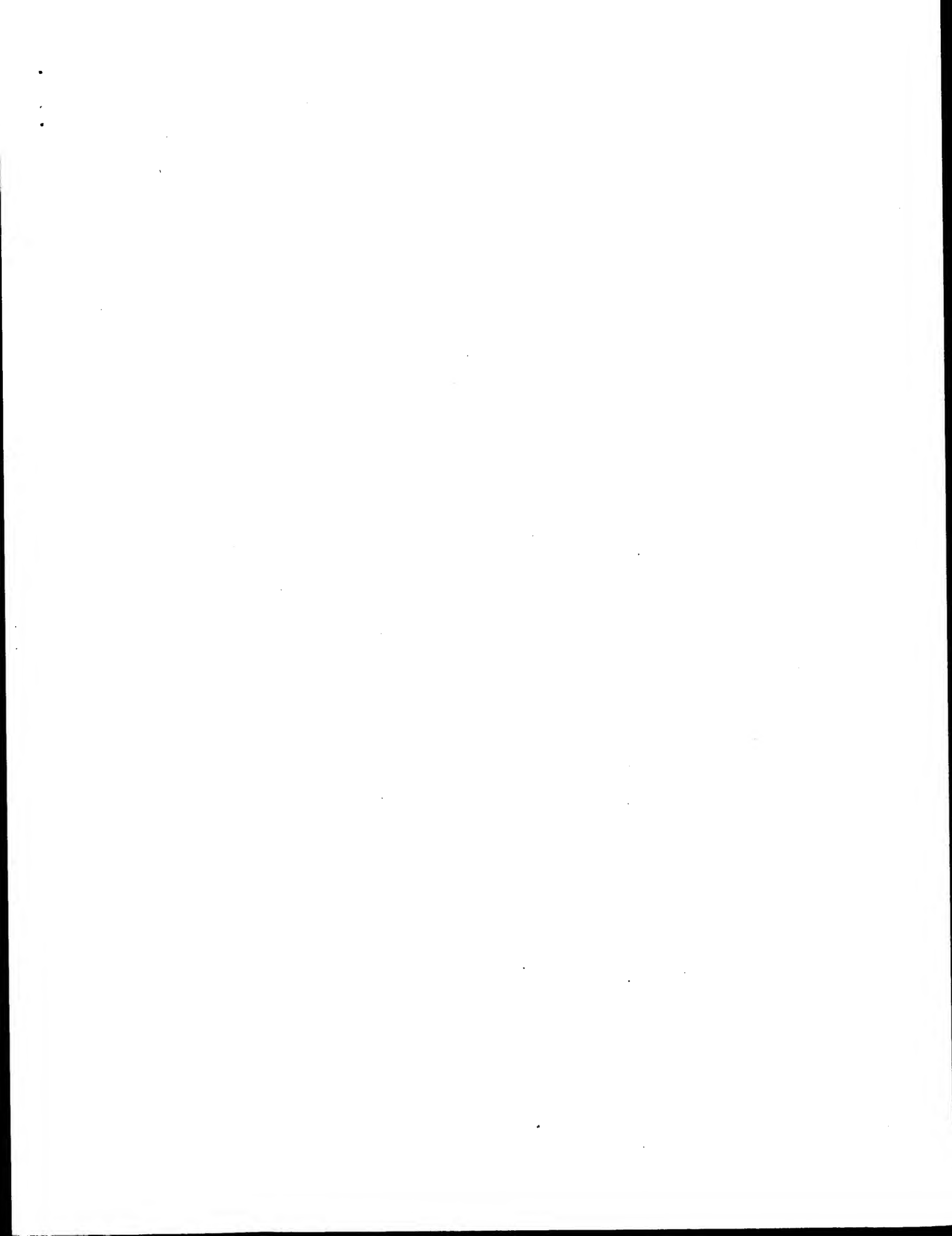
F:506-537/Domain: fungal cellulose-binding domain homology <FCB>

Query Match 11.0%; Score 183; DB 1; Length 537;

Best Local Similarity 23.4%; Pred. No. 1,1e-05;
Matches 79; Conservative 30; Mismatches 99; Indels 130; Gaps 14;

```
OY 24 GSGSTRYMDCKRPSKAMPKAAVSPYACDANPQRLSDFNVSGCNGSAYSADQTP 83
    | : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 268 GTVSTDRYGGTCDDP-----GCDENPRMGVTNRY----GPGETIDTRKSP 308
OY 84 WAV-----NDNLAYGFPA-----TSIAGS-ESSWCACAYALTF 116
    | : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 309 FTVVTQFLTNDGTSTGTLSEIKRPFYOGKVIQNPQSTIVSGNSITDSMCNA----- 362
OY 117 TSGPVAKTNVOSTSTGCLGSGNQFDI--AMPGGVGIENG----- 157
    | : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 363 -----OKSAFGD--TNEFSKHGGMAGLADGNVLVMSLWMDHASDMLML 407
OY 158 -----SSQFGGLPQAOYCGIISRDQDSFPAPLKPCCQWREDFQNDNPFETFOYQCP 212
    | : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 408 DSTYPTNATSTTGAK-----RCTDISRRP-----NTVESTY-----P 441
OY 213 AEIYARSGCKRNDSSFPVFTPPSGGNGGTGTPTSTAPGSGQ-----TSPGGS 261
    | : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 442 NAYVIYSNITGPLNS--FTTGGTSSSTTTTSTSTSSSKTWTVTYTTTTSSGSS 499
OY 262 GCTSOKWAOCGGIGFSGCTCVSGTTCQKLNDDYSSQL 299
    | : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 500 GTGARDMAOCGGNGMTGPTTCVSPYTCTKONDMYSOCL 537
```

Search completed: February 24, 2003, 11:52:59
Job time : 20 secs



Matches 183; Conservative 36; Mismatches 72; Indels 79; Gaps 6;

QY 9 TTAAALPLVASASGSGSTRYWCCKPSCAMPKKAIVSQVYACDANFORLSDFNVS 68
 DB 5 TLLALAGPLVASASGSHSTRYWCCKPSCMSGSAVANAALPTCDKNDPISNTNAYN 64
 QY 69 GC-NGGSAYSCADOPFMAVNDLAVGFATSIAGSESSWCCACVAFPTSGPVAKTMY 127
 DB 65 GEGGGSAYACTNYSFMAVNDLAVGFATSIAGSESSWCCACVAFPTSGPVAKTMY 124
 QY 128 VQSTGTGDLGSMOPDIAMPGGVGIFNGCSSQFG-GLPGAQYGGISSRDCCSFAPL 186
 DB 125 VQSTGTGDLGSMOPDIAMPGGVGIFNGCSSQFG-GLPGAQYGGISSRDCCSFAPL 184
 QY 187 PGCOMRFMFQVADNPTFTFOVQCPAEIVARSCKRNDSSPFV----- 232
 DB 185 DGCRRFMEFENADNPTFTFOVQCPALDLSGCKRNDSSPFAPKGDTSASKPOBSS 244
 QY 233 -----TPPS-----GNGGTGTPTSTAPSGQTS 257
 DB 245 AKKTTSAANAOPQTKDSAPVYQKSTKPAAPPEPTKPADKQDTKPAATKPAQ 304
 QY 238 GGGSGCTSQK-----NAQCGIGF---SGCTTCVSGTTCQ 289
 DB 305 PYNKPKTTQKVGTKTRGSCPAKTDATAKASVVPAYQCGSKSAYPNGLACATGKCV 364
 QY 290 KUNDYYSOCL 299
 DB 365 KONEYYSOCV 374

RESULT 2
 GUN5_HUMIN STANDARD: PRT: 213 AA.
 ID GUN5_HUMIN
 AC P43316;
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Endoglucanase V (EC 3.2.1.4) (Endo-1,4-beta-glucanase V)
 OS Humicola insolens.
 OS Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Humicola.
 OC NCBI_TaxID=34413;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Rasmussen G., Mikkelsen J.-M., Schulein M., Patkar S.A., Hagen F.,
 RA Hjort C.M., Hestrup S.;
 RT "A cellulase preparation comprising an endoglucanase enzyme.";
 RL Patent number WO9117243, 14-NOV-1991.
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).
 RX MEDLINE=93390621; Pubmed=8377830;
 RA Davies G.J., Dodson G.G., Hubbard R.E., Tolley S.P., Dauter Z.,
 RA Wilson K.S., Hjort C., Mikkelsen J.M., Rasmussen G., Schulein M.;
 RT "Structure and function of endoglucanase V.";
 RL Nature 365:362-364(1993).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
 RX MEDLINE=96101453; Pubmed=8519779;
 RA Davies G.J., Tolley S.P., Henriksat B., Hjort C., Schulein M.;
 RT "Structures of oligosaccharide-bound forms of the endoglucanase V
 from Humicola insolens at 1.9-A resolution.";
 RL Biochemistry 34:16210-16220(1995).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS).
 RA Davies G.J., Dodson G.G., Moore M.H., Tolley S.P., Dauter Z.,
 RA Wilson K.S., Rasmussen G., Schulein M.;
 RT "Structure determination and refinement of the Humicola insolens
 endoglucanase V at 1.5-A resolution.";
 RL Acta Crystallogr. D 52:7-17(1996).
 CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
 CC linkages in cellulose.
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY K (FAMILY 45 OF GLYCOSYL

CC HYDROLASES).
 DR PDB; 2ENG; 08-DEC-96.
 DR PDB; 3ENG; 16-JUN-97.
 DR PDB; 4ENG; 16-JUN-97.
 DR InterPro: IPR000334; GH_45.
 DR Pfam: PF02015; Glyco_hydro_45; 1.
 DR PROSITE: PS01140; GLYCOSYL_HYDROL_F45; 1.
 KW Cellulose degradation; Hydrolase; Glycosidase; 3D-structure.
 FT ACT_SITE 10 10 NUCLEOPHILE
 FT ACT_SITE 121 121 PROTON DONOR.
 SQ SEQUENCE 213 AA; 22864 MW; 24334301BA3BC804 CRC64;

Query Match 55.5%; Score 921.5; DB 1; Length 213;
 Best Local Similarity 74.4%; Pred. No. 5.8e-61;
 Matches 157; Conservative 27; Mismatches 26; Indels 1; Gaps 1;

QY 26 GOSTRYWDCCKPSCAMPKKAIVSQVYACDANFORLSDFNVSQGN-GGSAYSCADQTPW 84
 DB 3 GRSTRYWDCCKPSCAMPKKAIVSQVYACDANFORLSDFNVSQGN-GGSAYSCADQTPW 62
 QY 85 AVNDLAVGFATSIAGSESSWCCACVAFPTSGPVAKTMYVQSTGTGDLGSMOPDI 144
 DB 63 AVNDLAVGFATSIAGSESSWCCACVAFPTSGPVAKTMYVQSTGTGDLGSMOPDI 122
 QY 145 AMPGGVGIFNGCSSQFG-GLPGAQYGGISSRDCCSFAPLPGCOMRFMFQVADNPT 204
 DB 123 NIPGGVGIFNGCSSQFG-GLPGAQYGGISSRDCCSFAPLPGCOMRFMFQVADNPT 182
 QY 205 TFOVQCPAEIVARSCKRNDSSPFVFPMP 235
 DB 183 SFRVQCPAEIVARSCKRNDSSPFVFPMP 213

RESULT 3
 GUNB_PSEFL STANDARD: PRT: 511 AA.
 ID GUNB_PSEFL
 AC P18126;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Endoglucanase B precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
 OS CELB.
 OS Pseudomonas fluorescens.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 OC NCBI_TaxID=294;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 30-48.
 RC STRAIN-SP. Cellulosa;
 RX MEDLINE=9035836; Pubmed=2117693;
 RA Gilbert H.J., Hall J., Hazlewood G.P., Ferreira L.M.A.;
 RT "The N-terminal region of an endoglucanase from Pseudomonas
 fluorescens subspecies cellulosa constitutes a cellulose-binding
 domain that is distinct from the catalytic centre.";
 RL Mol. Microbiol. 4:759-767(1990).
 CC -1- FUNCTION: THIS ENZYME CATALYZES THE ENDOHYDROLYSIS OF 1,4-BETA-
 CC GLUCOSIDIC LINKAGES IN CELLULOSE, LICHENIN AND CEREAL BETA-D-
 CC GLUCANS. EGB IS MOST ACTIVE AGAINST BARLEY BETA-GLUCAN, BUT SHOWED
 CC SIGNIFICANT ACTIVITY AGAINST AMORPHOUS AND CRYSTALLINE CELLULOSE.
 CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
 CC linkages in cellulose.
 CC -1- SUBCELLULAR LOCATION: Periplasmic.
 CC -1- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN
 CC (CBD).
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY K (FAMILY 45 OF GLYCOSYL
 CC HYDROLASES).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; X52615; CAA36844.1; -
 CC PIR; S10527; S10527.
 DR HSSP; P43316; 2ENG.
 DR InterPro; IPR001919; Bac_cellose-blnd.
 DR InterPro; IPR002883; CBD_5.
 DR Pfam; PF00553; CBM_2; 1.
 DR Pfam; PF02013; CBM_10; 1.
 DR Pfam; PF02015; Glyco_hydro_45; 1.
 DR PROSITE; PS00561; CBD_BACTERIAL; 1.
 DR PROSITE; PS01140; GLYCOSYL_HYDROL_F45; 1.
 KW Cellulose degradation; Hydrolase; Glycosidase; Signal; Periplasmic.
 FT SIGNAL 1 29
 FT CHAIN 30 511
 FT DOMAIN 30 131
 FT DOMAIN 132 173
 FT DOMAIN 223 259
 FT DISULFID 32 127
 FT ACT_SITE 276 276
 FT ACT_SITE 393 393
 FT ACT_SITE 393 393
 SO SEQUENCE 511 AA; 52078 MW; 3C3119D998291D8E CRC64;

Query Match 30.6%; Score 507; DB 1; Length 511;
 Best Local Similarity 39.6%; Pred. No. 3.1e-30;
 Matches 107; Conservative 38; Mismatches 73; Indels 52; Gaps 7;

OY 2 RSPVRLTTLAALPLVSAASG-----SGQSTRYWDCCPSCAMPCKA-AVSOPVY 52
 DB 237 RSVVASSSSSLSSTSSASSVSPPIIDGGNGATRYWDCCPSCGMSANVPSLVSPLO 296
 OY 53 ACDANFORLDFENVQSGNCGSAYSCADQTPMAVNDNLAYFAFATSIAGSESSMCCACY 112
 DB 297 SCGANRTLRSDVSVSGSCGGGCMGDKTFPAVSPTLACVAT-----SSDVCGRCY 351
 OY 113 ALFTT-----SGPVAKTMYVOSTSTGGLGSGNOFDIAMPGGVGIENGSSOP 161
 DB 352 QLOFTSSSYNAPDPPGSAALAGKTMTVOATNIGDYSGGFDILVPGGVGAGVACNASQW 411
 OY 162 G---GLPGAQYGGISSR-----DQDQSF-----PAFLKGCQCMRPD 194
 DB 412 GYSNMLGAGYGFGLACQOGLYNASLSQYSVCULNRCDVSFGSGRLQLOQGTWFAE 471
 OY 195 WFOQADNPTFTFOQVCPAEIVARSGCKRN 224
 DB 472 WFEADNPSLKYKEVCPAELTTRSGMNR 501

RESULT 4
 GUN1_USTMA STANDARD; PRT; 393 AA.
 AC P54424;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Endoglucanase I precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase 1)
 DE (Cellulase 1) (EG 1).
 GN EglI.
 OS Ustilago maydis (Smut fungus).
 CC Eukaryota; Fungi; Basidiomycota; Ustilaginomycetes;
 CC Ustilaginomycetidae; Ustilaginales; Ustilaginaceae; Ustilago.
 OX NCBI_TaxID=5270;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=FB11;
 RX MEDLINE=96145728; PubMed=8590631;
 RA Schaeucker F., Wanner G., Kahmann R.;
 RT "Filament-specific expression of a cellulase gene in the dimorphic
 fungus Ustilago maydis.";
 RL Biol. Chem. Hoppe-Seyler 376:617-625(1995).
 CC -1- CATALYTIC ACTIVITY: Endohydrolase of 1,4-beta-D-glucosidic

CC linkages in cellulose.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- TISSUE SPECIFICITY: HYPHAL TIP.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN FILAMENTOUS DIKARYON.
 CC -1- PTM: MAY ALSO BE O-GLYCOSYLATED.
 CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY K (FAMILY 45 OF GLYCOSYL
 CC HYDROLASES).

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; S81598; AAB36147.1; -
 DR HSSP; P43316; 2ENG.
 DR InterPro; IPR000334; GH_45.
 DR Pfam; PF02015; Glyco_hydro_45; 1.
 DR PROSITE; PS01140; GLYCOSYL_HYDROL_F45; 1.
 KW Cellulose degradation; Hydrolase; Glycosidase; Signal; Glycoprotein.
 FT SIGNAL 1 26
 FT CHAIN 27 393
 FT ACT_SITE 34 34
 FT ACT_SITE 152 152
 FT DOMAIN 270 385
 FT CARBOHYD 343 343
 SO SEQUENCE 393 AA; 39594 MW; 65C753C610CD6AD3 CRC64;

Query Match 27.8%; Score 462; DB 1; Length 393;
 Best Local Similarity 37.7%; Pred. No. 4.7e-27;
 Matches 117; Conservative 41; Mismatches 106; Indels 46; Gaps 16;

OY 11 LAALPLVSAASGSGSTRYWDCCPSCAMPCKAAYVOPVYACDANFORL-----SPFN 66
 DB 12 LSLSLVHLDGYRAGATRYWDCCLASASMEKAPYAPVADKADGVTLLDSKKP 71
 OY 67 QSGCNGSAYSCADQTPW--AVNDNLAYFAFATSIAGSESSMCCACYALTFTSGPVAK 124
 DB 72 QSGCNGSNKFMSCMOFPDEPTPLAFGGA--FTTGQSSDPRDCACFYAEF-EHDAQK 128
 OY 125 TM-----VVOSTSTGGLGSGNOFDIAMPGGVGIENGSSOPG---GLPGAQYGGISSR 175
 DB 129 AMKRNKLIFQVTWVGDDVQSNDFQIPGGGLGAFPRGCPAOMGVESLMDGYGVKSA 188
 OY 176 DQDQSPAPLKPCCQWR--DFQONADNPTF--TFOQVCPAEIVARSGCKRNDSSFPVF 232
 DB 189 TECSKLKPLQEGCKMRFSEM--GDNVPLKGSPPKRYCKSLIDRSGCQRKDDNT---I 242
 OY 233 TPSPG--GNGGTGTPTS-----TAPG--SGQTSPPG--GSGCTSQKMAQCGGIGFSGC 279
 DB 243 SPYSGKVDNANTPAPOYKRDNRVCLLAGKKKSAAGAGVDSGDAS-----GGAASA 296
 OY 280 TTCVSGCTCQ 289
 DB 297 GGAAGSGQ 306

RESULT 5
 GUN5_TRIE STANDARD; PRT; 242 AA.
 AC P43317;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Endoglucanase V precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase V)
 DE (Cellulase V) (EG V).
 GN Egl5.
 OS Trichoderma reesei (Hypocrea jecorina).
 CC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 CC Hypocreales; Hypocreaceae; Hypocrea.
 OX NCBI_TaxID=51453;

[1]
 RN SEQUENCE FROM N.A.
 RP STRAIN-OM9414 / Rut C-30; PubMed=7984103;
 RC MEDLINE=95075308; PubMed=7984103;
 RA Saloheimo A., Henttinen B., Henttinen A.-M., Telemann O., Penttilä M.;
 RT "A novel, small endoglucanase gene, eg15, from *Trichoderma reesei*
 isolated by expression in yeast."
 RL Mol. Microbiol. 13:219-228(1994).
 CC -1. CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
 linkages in cellulose.
 CC -1. SIMILARITY: BELONGS TO CELLULOSE FAMILY K (FAMILY 45 OF GLYCOSYL
 HYDROLASES).
 CC -1. SIMILARITY: CONTRAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z33381; CAA83846.1; -.
 DR HSSP: P00725; 2CBH.
 DR InterPro: IPR000334; GH_45.
 DR Pfam: PF00734; CBM_1; 1.
 DR Pfam: PF02015; Glyco_hydro_45; 1.
 DR Prodom: PD001821; CBD_fungal; 1.
 DR SMART: SM00236; fCBD; 1.
 DR PROSITE: PS00562; CBD_FUNGAL; 1.
 DR PROSITE: PS00842; EXPANSIN_EG45; 1.
 DR PROSITE: PS01140; GLYCOSTYL_HYDROL_45; 1.
 DR Cellulose degradation; Hydrolyase; signal.
 DR KW SIGNAL 1 17
 FT CHAIN 18 242
 FT ACT_SITE 27 27
 FT ACT_SITE 134 134
 FT DOMAIN 18 182
 FT DOMAIN 183 205
 FT DOMAIN 206 242
 FT CAROXYD 182 182
 FT DISULFID 213 230
 FT DISULFID 224 240
 SEQUENCE 242 AA; 2441 MW; CC033FC51326C71D CRC64;
 Query Match 13.3%; Score 221; DB 1; Length 242;
 Best Local Similarity 24.7%; Pred. No. 1.3e-09;
 Matches 78; Conservative 39; Mismatches 101; Indels 98; Gaps 13;
 OY 7 LRTTLAALPLVVASAGSGSPTRYMD-----CCKPSCAMPKKAIVSQVYACDANFQ 59
 DB 1 MKATLVAGSLIVGAVSAVKATTTTRTYDQEGACGCCSSGSAFPQOLGIGNVYTA-AASQ 59
 OY 60 RLSDPFWQSGCNGSGAYSACDQTPWAVNDMLAVGAFTSIAGSESSWCACALYFTF-- 117
 DB 60 ALPDTGASGAG--G-----CGKCYQLTSTGQ 85
 OY 118 -----SGPVAGRTMAYVSTSTGGDLGNSNFDIAMPGGGVGIRNGSSQGGGLPAGAYG 170
 DB 86 APCSSCGTGAAGOSITVMTNLCPNNGNMAQCPV--GGTN-----OY- 127
 OY 171 GISSRQDCDFPAPLKGCCWRPD-WFON--ADNPFTFOVOCFAIVARSG---CKR 223
 DB 128 -----GYSTHFDIIMANETFGDNVYVDFEPIACPGQASWGTCLCYG 170
 OY 224 NDDSPFVFTPPSGGNGGTPTSTAPSGQTSPOGSGCTSQKWAQCGGIGFSCTTCV 283
 DB 171 QQETD-P--TPVLGNDGSGTPPGSSPATSSPSGGGQGT--LYGCGGAGWGTPTTCQ 225
 OY 284 SGTTCOKLNDIYSOCL 299
 DB 226 APGTCKVONOMYSOCL 241

RESULT 6
 GUX1_TRIRE
 ID GUX1_TRIRE STANDARD: PRT; 513 AA.
 AC P00725;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Exoglucanase I precursor (EC 3.2.1.91) (Exocellulohydrolase I)
 DE (CBH1) (1,4-beta-cellulohydrolase).
 GN CBH1.
 OS *Trichoderma reesei* (Hypocrea jecorina), and
 OS *Trichoderma konigii*.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Hypocreales; Hypocreaceae; Hypocrea.
 OX NCBI_TaxID=51453; 55202;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES-T. reesei; STRAIN=L27;
 RA Shewmaker S., Schweickart V., Ladner M., Gelfand D., Kwok S.,
 RT "Molecular cloning of exo-cellulohydrolase I derived from *Trichoderma*
 reesei strain L27."
 RL Biotechnology 1:691-696(1983).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES-T. konigii; STRAIN=G-39;
 RA MEDLINE=94100788; PubMed=7764306;
 RT Wey T.T., Hsu T.H., Huang L.;
 RT "Molecular cloning and sequence analysis of the cellulohydrolase I
 gene from *Trichoderma konigii* G-39."
 RL Curr. Microbiol. 28:31-39(1994).
 RN [3]
 RP ACTIVE SITE.
 RC SPECIES-T. reesei;
 RA Tomme P., Claysens M.;
 RT "Identification of a functionally important carboxyl group in
 cellulohydrolase I from *Trichoderma reesei*."
 RL FEBS Lett. 243:239-243(1989).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (1.81 ANGSTROMS) OF 18-452.
 RA MEDLINE=94310436; PubMed=8036495;
 RT Divne C., Stahlberg J., Teeri T.T., Ruohonen L., Pettersson G.,
 RA Knowles J.K.C., Teeri T.T., Jones T.A.;
 RT "The three-dimensional crystal structure of the catalytic core of
 cellulohydrolase I from *Trichoderma reesei*."
 RL Science 265:524-528(1994).
 RN [5]
 RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS) OF 18-452.
 RC SPECIES-T. reesei; STRAIN-OM9414 / Rut C-30;
 RA MEDLINE=98128795; PubMed=9466911;
 RT Divne C., Stahlberg J., Teeri T.T., Jones T.A.;
 RT "High-resolution crystal structures reveal how a cellulose chain is
 bound in the 50 Å long tunnel of cellulohydrolase I from *Trichoderma*
 reesei."
 RL J. Mol. Biol. 275:309-325(1998).
 RN [6]
 RP STRUCTURE BY NMR OF 478-513.
 RC SPECIES-T. reesei;
 RA MEDLINE=90057416; PubMed=2554967;
 RT Kravits J.J., Clore G.M., Nilges M., Jones T.A., Pettersson G.,
 RA Knowles J.J., Gronenborn A.M.;
 RT "Determination of the three-dimensional solution structure of the C-
 terminal domain of cellulohydrolase I from *Trichoderma reesei*. A
 study using nuclear magnetic resonance and hybrid distance
 geometry-dynamical simulated annealing."
 RL Biochemistry 28:7241-7257(1989).
 RN [7]
 RP STRUCTURE BY NMR OF 478-513.
 RC SPECIES-T. reesei;
 RA MEDLINE=97194052; PubMed=9041630;
 RT Mattinen M.L., Kontteli M., Kerovuo J., Linder M., Annala A.,

[illegible]

```

CC (2) EXOCELLULOBIODIHYDROLASES THAT CUT THE DISSACCHARIDE CELLULOSE
CC FROM THE NONREDUCING END OF THE CELLULOSE POLYMER CHAIN:
CC (3) BETA-1,4-GLUCOSIDASES WHICH HYDROLYZE THE CELLOBIOSE AND OTHER
CC SHORT CELLO-OLIGOSACCHARIDES TO GLUCOSE.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
CC in cellulose and cellobiose, releasing cellobiose from the non-
CC reducing ends of the chains.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY C (FAMILY 7 OF GLYCOSYL
CC HYDROLASES).
CC -1- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X53931; CAA37878.1; -.
CC PIR: S11439; S11439.
CC DR HSSP: P00723; BCEL.
CC DR InterPro: IPR000234; CBD_fungal.
CC DR InterPro: IPR001722; GH_7.
CC DR Pfam: PF00734; CBM_1; 1.
CC DR Pfam: PF00840; Glyco_hydro_7; 1.
CC DR ProDom: PD001821; CBD_fungal; 1.
CC DR ProDom: PD186135; GH_7; 1.
CC DR SMART: SM00236; fCBD; 1.
CC DR PROSITE: PS00562; CBD_FUNGAL; 1.
CC DR Cellulose degradation; Hydrolase; Glycosidase; Glycoprotein; Signal.
CC KW CHAIN 1 17
CC FT CHAIN 18 513 EXOGLUCANASE I.
CC FT DOMAIN 18 477 CATALYTIC.
CC FT DOMAIN 454 477 LINKER.
CC FT DOMAIN 478 513 CELLULOSE-BINDING (BY SIMILARITY).
CC FT ACT_SITE 229 229 NUCLEOPHILE (BY SIMILARITY).
CC FT ACT_SITE 234 234 PROTON DONOR (BY SIMILARITY).
CC FT CARBOHYD 62 62 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 81 81 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 287 287 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT DISULFID 485 502 BY SIMILARITY.
CC FT DISULFID 496 512 BY SIMILARITY.
CC SQ SEQUENCE 513 AA; 53891 MW; 58EF552717C4C8E CRC64;

Query Match 13.0%; Score 215.5; DB 1; Length 513;
Best Local Similarity 26.2%; Pred. No. 6.8e-09;
Matches 86; Conservative 36; Mismatches 137; Indels 69; Gaps 16;

```

```

CC (1) EXOCELLULOBIODIHYDROLASES THAT CUT THE DISSACCHARIDE CELLULOSE
CC FROM THE NONREDUCING END OF THE CELLULOSE POLYMER CHAIN:
CC (3) BETA-1,4-GLUCOSIDASES WHICH HYDROLYZE THE CELLOBIOSE AND OTHER
CC SHORT CELLO-OLIGOSACCHARIDES TO GLUCOSE.
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY C (FAMILY 7 OF GLYCOSYL
CC HYDROLASES).
CC -1- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M15665; AAA34212.1; -.
CC PIR: A25928; A25928.
CC PIR: A25655; A25655.
CC PDB: 1EG1; 19-MAR-99.

RESULT 8
GNL TRIRE
ID GUNL_TRIRE STRAND: PRT; 459 AA.
AC P07981;
DR 01-AUG-1988 (Rel. 08, Created)
DR 15-AUG-1988 (Rel. 08, Last sequence update)
DR 15-JUL-1999 (Rel. 38, Last annotation update)
DE Endoglucanase Eg-1 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE (Cellulase).
GN EglI.
OS Trichoderma reesei (Hypocrea jecorina).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreales; Hypocreaceae; Hypocrea.
OX NCBI_TaxID=51453;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-VTT-D-80133;
RX MEDLINE=87106822; PubMed=2948877;
RA Penttilae M., Lehtovaara P., Nevalainen H., Bhikhabhai R.,
RA Knowles J.K.C.;
RT "Homology between cellulase genes of Trichoderma reesei: complete
RT nucleotide sequence of the endoglucanase I gene.";
RL Gene 45:253-265(1986).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=L27;
RA van Arsdell J.N., Kwok S., Schweickart V.L., Ladner M.B.,
RA Gelfand D.H., Innis M.A.;
RT "Cloning, characterization, and expression in Saccharomyces cerevisiae
RT of endoglucanase I from Trichoderma reesei.";
RL Biotechnology 5:60-64(1987).
RN [3]
RP ACTIVE SITE GLU-149.
RA Tome P., Clavessens M.;
RT "Identification of a functionally important carboxyl group in
RT cellobiohydrolase I from Trichoderma reesei.";
RL FEBS Lett. 243:239-245(1989).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (3.6 ANGSTROMS) OF 23-393.
RX MEDLINE=97467423; PubMed=9325098;
RA Kleywegt G.J., Zou J.-Y., Dwyer C., Davies G.J., Slinning I.,
RA Staehelin J., Reinkainen T., Srisoduk M., Teeri T.T., Jones T.A.;
RT "The crystal structure of the catalytic core domain of endoglucanase
RT I from Trichoderma reesei at 3.6-A resolution, and a comparison with
RT related enzymes.";
RL J. Mol. Biol. 272:383-397(1997).
CC -1- FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE
CC GENERALLY REQUIRES THREE TYPES OF HYDROLYTIC ENZYMES:
CC (1) ENDOGLUCANASES WHICH CUT INTERNAL BETA-1,4-GLUCOSIDIC BONDS;
CC (2) EXOCELLULOBIODIHYDROLASES THAT CUT THE DISSACCHARIDE CELLULOSE
CC FROM THE NONREDUCING END OF THE CELLULOSE POLYMER CHAIN;
CC (3) BETA-1,4-GLUCOSIDASES WHICH HYDROLYZE THE CELLOBIOSE AND OTHER
CC SHORT CELLO-OLIGOSACCHARIDES TO GLUCOSE.
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY C (FAMILY 7 OF GLYCOSYL
CC HYDROLASES).
CC -1- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M15665; AAA34212.1; -.
CC PIR: A25928; A25928.
CC PIR: A25655; A25655.
CC PDB: 1EG1; 19-MAR-99.

```


DR InterPro: IPR000254; CBD_fungal.
 DR InterPro: IPR001722; GH_7.
 DR Pfam: PF00734; CBM_1; 1.
 DR Pfam: PF00840; Glyco_hydro_7; 1.
 DR Prodom: PD186135; GH_7; 1.
 DR SMART: SM00236; fCBD; 1.
 DR PROSITE: PS00562; CBD_FUNGAL; 1.
 DR Cellulose degradation; Hydrolase; glycosidase; glycoprotein; signal;
 3D-structure.
 FT SIGNAL 1 22
 FT CHAIN 23 459
 FT DOMAIN 23 397
 FT DOMAIN 398 423
 FT DOMAIN 424 459
 FT ACT_SITE 149 149
 FT ACT_SITE 218 218
 FT ACT_SITE 223 223
 FT CARBOHYD 78 78
 FT CARBOHYD 164 164
 FT CARBOHYD 204 204
 FT CARBOHYD 208 208
 FT CARBOHYD 394 394
 FT DISULFID 431 448
 FT DISULFID 442 458
 FT SEQUENCE 459 AA; 48208 MW; D235A256F808CB9 CRC64;
 12.9%; Score 214; DB 1; Length 459;
 Best local similarity 27.0%; Pred. No. 7.9e-09;
 Matches 108; Conservative 31; Mismatches 101; Indels 160; Gaps 27;

QY 19 ASASGSGS---TRYWDCCKPSCAMPKAAVSOPIYACDA-----NQRLSPFN 66
 DB 101 ASGVTTSGSLTWNQ---MPSS--GGYSSVSPRLYLDSDEYMLKNGELS-FDV 154
 QY 67 Q-----SGCGSASGADPT--WAVNDNLAYGFA 95
 DB 155 DLSALPCGSLYLSQMDENGANGANYGTAGANYGSGY-CDACPVQYWR-NGTL----- 207
 QY 96 AVSIAGSESSMCC-----AC-----VALFTFS--GPV 121
 DB 208 -----NTSHQGCNEMDILEGSRANALTPHSCRTACDSAGCGNPGSGKSYGP- 261
 QY 122 ACKTMYVOST-----STGGDLGS-----NOFDI--AMPGGVGIFNGCSSQ 160
 DB 262 -GDVDTSTKFTIITQFNTDNGSPSGNLVSTRKYQNGVDIPSAQPGDT--ISSCP- 317
 QY 161 FGLGLGAOYGGISSRQCCSFAPLPKPGQWRD-WFQAND--NPTFTTQOVCC----- 211
 DB 318 -----ASAYGLATWGKA-----LSSGMVLVFSIWNDSQYWNMLDSGNAGPCSTEGN 366
 QY 212 PAEIVARSGCKRNDSSFPVFTPPSGNGGTGTPRTA---PGSGQT-----SPGG 259
 DB 367 PSNILA-----NPNPTHVFSNIRMGDLS--TTNSTAPPPPASSTTSTRRSSTSS 419
 QY 260 GSGCTISQKAAQCGGIGFSCCTCVSGTTCOKLNDYISGL 299
 DB 420 SPSCQTQTHWGCGIGCYGSGCKTCTSGTTCQYSDYSSQCL 459

RESULT 9
 GUNI_TRILLO STANDARD: PRT: 463 AA.
 AC Q12714:
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Endoglucanase Eg-1 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
 DE (Cellulase).
 DE EGL1.
 OS Trichoderma longibrachiatum.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Hypocreales; mitosporic Hypocreales; Trichoderma.
 NCBI_TaxID=5548;

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CECT 2606;
 RA Perez-Gonzalez J.A.;
 RL Submitted (NOV-1992) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE
 CC GENERALLY REQUIRES THREE TYPES OF HYDROLYTIC ENZYMES:
 CC (1) ENDOGLUCANASES WHICH CUT INTERNAL BETA-1,4-GLUCOSIDIC BONDS;
 CC (2) EXOCELLULOBIOSYDOLASES THAT CUT THE DISSACCHARIDE CELLULOSE
 CC FROM THE NONREDUCING END OF THE CELLULOSE POLYMER CHAIN;
 CC (3) BETA-1,4-GLUCOSIDASES WHICH HYDROLYZE THE CELLULOSE AND OTHER
 CC SHORT CELLO-OLIGOSACCHARIDES TO GLUCOSE.
 CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
 CC linkages in cellulose.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY C (FAMILY 7 OF GLYCOSYL
 CC HYDROLASES).
 CC -1- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb.slb.ch).

DR EMBL: X60652; CAA43059.1; -.
 DR HSSP: P07981; IEGL.
 DR InterPro: IPR000254; CBD_fungal.
 DR InterPro: IPR001722; GH_7.
 DR Pfam: PF00734; CBM_1; 1.
 DR Pfam: PF00840; Glyco_hydro_7; 1.
 DR Prodom: PD186135; GH_7; 1.
 DR SMART: SM00236; fCBD; 1.
 DR PROSITE: PS00562; CBD_FUNGAL; 1.
 DR Cellulose degradation; Hydrolase; glycosidase; glycoprotein; signal.
 FT SIGNAL 1 22
 FT CHAIN 23 463
 FT DOMAIN 23 397
 FT DOMAIN 402 427
 FT DOMAIN 428 463
 FT ACT_SITE 218 218
 FT ACT_SITE 223 223
 FT ACT_SITE 435 452
 FT DISULFID 446 462
 FT CARBOHYD 78 78
 FT CARBOHYD 164 164
 FT CARBOHYD 204 204
 FT CARBOHYD 208 208
 FT CARBOHYD 394 394
 FT SEQUENCE 463 AA; 48337 MW; B3AC3DPD3ADD2B1C CRC64;
 12.2%; Score 203; DB 1; Length 463;
 Best local similarity 28.4%; Pred. No. 5.1e-08;
 Matches 86; Conservative 22; Mismatches 101; Indels 94; Gaps 18;

QY 68 SGCNGSASGADPT--WAVNDNL--AYGPAAT--SIAGSESSMCCACALFTSG 119
 DB 184 AGANYSGY-CDACPVQYWR-NGTLNTSGGFCNEMDILEGSRANALTPHSCYATAC 241
 QY 120 PVAG-----KTMVY-----OSTSGDLGS-----NOFD 143
 DB 242 DSAAGCFNRYGSGYRNYRFGDVTDSKFTITTOFNTDNGSSGNLVSTRKYRQNGVD 301
 QY 144 I--AMPGGVGIFNGCSSQFGLGAOYGGISSRQCCSFAPLPKPGQWRD-WFQAND 200
 DB 302 IPSAKPGDT--ISSCP-----ASAYGLATWGKA-----LSSGMVLVFSIWNDSQ 347
 QY 201 -----NPTFTTQOVCPAEIVARSGCKRNDSSFPVFTPPSGN--- 239
 DB 348 YWNMLDSGRAGPCSTEGNPSNLT--ANNPPTHVFSNIRMGDLS--TTNSTGPNPP 402

OY 240 ---GGCTPTSTAPSGGOTSPPGGSGCTSQKNAOCGGISGCTTCVSTTCCKINDYYS 296
 Db 403 PPPPASSTFTSTFRSSSTYS--SSPSCOTPHMGCGGIGTGTCTGTTCGTTCGTGNDYYS 460
 OY 297 OCL 299
 |||
 Db 461 OCL 463

RESULT 10
 GUX1_HUMGR STANDARD: PRT; 525 AA.
 ID GUX1_HUMGR
 AC P15828:
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Exoglucanase I precursor (EC 3.2.1.91) (Exocellulobiohydrolase I) (1,4-beta-cellobiohydrolase) (Beta-glucanocellulobiohydrolase).
 CN CBH-1.
 OS Humicola grisea.
 OC Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Humicola.
 OX NCBI_TaxID=5527;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=var. thermolidea;
 RX MEDLINE=90175006; PubMed=2308855;
 RA de Oliveira Alzevedo M., Radford A.;
 RT "Sequence of cbh-1 gene of Humicola
 RL Nucleic Acids Res. 18:668-668(1990).
 CC -1- FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE
 CC GENERALLY REQUIRES THREE TYPES OF HYDROLYTIC ENZYMES:
 CC (1) ENDOGLUCANASES WHICH CUT INTERNAL BETA-1,4-GLUCOSIDIC BONDS;
 CC (2) EXOCELLULOBIHYDROLASES THAT CUT THE DISSACCHARIDE CELLULOSE
 CC FROM THE NONREDUCING END OF THE CELLULOSE POLYMER CHAIN;
 CC (3) BETA-1,4-GLUCOSIDASES WHICH HYDROLYZE THE CELLULOSE AND OTHER
 CC SHORT CELLO-OLIGOSACCHARIDES TO GLUCOSE.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
 CC in cellulose and cellotetraose, releasing cellobiose from the non-
 CC reducing ends of the chains.
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY C (FAMILY 7 OF GLYCOSYL
 CC HYDROLASES).
 CC -1- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X17258; CAA35159.1; -
 CC PIR: S08240; S08240.
 CC HSSP: P00725; 8CEL.
 CC InterPro: IPR000254; CBD_fungal.
 CC InterPro: IPR001722; GH_7.
 CC Pfam: PF00734; CBM_1.1.
 CC Pfam: PF00840; Glyco_hydro_7.1.
 CC ProDom: PD001821; CBD_fungal.1.
 CC ProDom: PD186135; GH_7.1.
 CC SMART: SM00236; fCBD.1.
 CC PROSITE: PS00562; CBD_FUNGAL.1.
 CC KMW Cellulose degradation; Hydrolase; Glycosidase; Glycoprotein; Signal.
 FT SIGNAL 1 18
 FT CHAIN 19 525
 FT DOMAIN 19 467
 FT DOMAIN 468 489
 FT DOMAIN 490 525
 FT ACT_SITE 231
 FT ACT_SITE 236
 FT CARBOHYD 289
 FT DISULFID 497
 FT DISULFID 508

SO SEQUENCE 525 AA; 55693 MW; A6684D4CF8B1E090 CRC64;
 Query Match 11.2%; Score 185; DB 1; Length 525;
 Best Local Similarity 24.9%; Pred. No. 1,2e-06;
 Matches 97; Conservative 33; Mismatches 136; Indels 124; Gaps 23;

OY 11 LAALPLVASAASGSGSTRYWDCKPSCAMPG-KAASOPVYACDANFORLSDP----- 64
 Db 159 LAGALYFVSMADGG-LSRY-----PGNKAGARYGTGCDACQPRDLKFTNGEA 206
 OY 65 NVQSGCN-----GGSAYS--CADQTPAAVNDNLAYGFA--ATSIAGS--ESSWCAC 111
 Db 207 NIEGWTGSTNDPNMAGAGRGTCGSEMDIWEAN-NMATAFPHPCITLIGSRCEGDCGGT 265
 OY 112 YALFTSG-----PVAKTYVYST-----STGDLGS-NDF 142
 Db 266 YSNERAYGVCDDGCDENFYRGCKNTFYKGMVTDTTKITVYTFQLKANGDGLGEIKRF 325
 OY 143 -----DIAMPG--GGVGIENGSSQ--FGGLPG-AOYGGISRHQDCDFPAPL 185
 Db 326 YVQDGKILPNSSTIPGVGNSITQDMCDROKVAFGDIDDFNKKGMKQWKALAGPMVL 385
 OY 186 KPGCGRPEDFQMANDPTE---TFQVOCPAELVARSGCKRN---DSSPPV----- 231
 Db 386 VMSI-----WDDHASNMMLDSTF-----PVDAAGPAGERGACPTTSGVPAEVEAPN 435
 OY 222 -----FTP-----PSGGNGTGTG---TPSTAPSGGOTSPPGGSGCTSQKMA 269
 Db 436 SNVVSNIREFGRTIGSTVVALPFGANGGNGNNGNPPTTTTSSAPATTTTASAPKGRWQ 495
 OY 270 OCGGIGFSGCTTCVSTTCCKINDYYSQCL 299
 Db 496 OCGGIGFTPTQCEEPYICTKINDYYSQCL 525

RESULT 11
 GUX1_PHACH STANDARD: PRT; 516 AA.
 ID GUX1_PHACH
 AC P13860;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Exoglucanase I precursor (EC 3.2.1.91) (Exocellulobiohydrolase I)
 DE (1,4-beta-cellobiohydrolase).
 GN CBH1.
 OS Phanerochaete chrysosporium.
 OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
 OC Aphelphorales; Corticiaceae; Phanerochaete.
 OX NCBI_TaxID=5306;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MEA46;
 RC MEDLINE=89232732; PubMed=3246351;
 RX Sims P.F.G., James C., Broda P.;
 RA "The identification, molecular cloning and characterisation of a gene
 RT from Phanerochaete chrysosporium that shows strong homology to the
 RT exo-cellobiohydrolase I gene from Trichoderma reesei.";
 RL gene 74:411-422(1988).
 RP [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MEA46;
 RC MEDLINE=9435641; PubMed=8057846;
 RX Sims P.F.G., Soares-Felipe M.S., Wang Q., Gent M.E., Tempelaar C.,
 RA Broda P.;
 RA "Differential expression of multiple exo-cellobiohydrolase I-like
 RT genes in the lignin-degrading fungus Phanerochaete chrysosporium.";
 RL Mol. Microbiol. 12:209-216(1994).
 CC -1- FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE
 CC GENERALLY REQUIRES THREE TYPES OF HYDROLYTIC ENZYMES:
 CC (1) ENDOGLUCANASES WHICH CUT INTERNAL BETA-1,4-GLUCOSIDIC BONDS;
 CC (2) EXOCELLULOBIHYDROLASES THAT CUT THE DISSACCHARIDE CELLULOSE
 CC FROM THE NONREDUCING END OF THE CELLULOSE POLYMER CHAIN;
 CC (3) BETA-1,4-GLUCOSIDASES WHICH HYDROLYZE THE CELLULOSE AND OTHER

CC SHORT CELLO-OLIGOSACCHARIDES TO GLUCOSE.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
CC in cellulose and cellobiose, releasing cellobiose from the non-
CC reducing ends of the chains.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY C (FAMILY 7 OF GLYCOSYL
CC HYDROLASES).
CC -1- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M22220; AAB46373.1; -;
DR EMBL: Z22528; CAA80253.1; -;
DR PIR: J50083; J50083.
DR HSSP: P00725; 2CBH.
DR InterPro: IPR000254; CBD_fungal.
DR InterPro: IPR001722; GH_7.
DR Pfam: PF00734; CBM_1; 1.
DR Pfam: PF00840; Glyco_hydro_7; 1.
DR ProDom: PD001821; CBD_fungal; 1.
DR ProDom: PD186135; GH_7; 1.
DR SMART: SM00236; fCBD; 1.
DR PROSITE: PS00562; CBD_FUNGAL; 1.
KM Cellulose degradation; Hydrolyase; glycoprotein; signal.
FT SIGNAL 1
FT CHAIN 19 516
FT DOMAIN 7 449
FT DOMAIN 450 480
FT DOMAIN 481 516
FT ACT_SITE 225 225
FT ACT_SITE 230 230
FT CARBOHYD 208 208
FT CARBOHYD 326 326
FT CARBOHYD 442 442
FT DISULFID 488 505
FT DISULFID 499 515
FT CONFLICT 27 28
FT CONFLICT 30 31
SQ SEQUENCE 516 AA; 54857 MW; 1C7C3D338ECE1B72 CRC64;

Query Match 11.1%; Score 184; DB 1; Length 516;
Best Local Similarity 24.7%; Pred. No. 1,4e-06;
Matches 86; Conservative 37; Mismatches 113; Indels 112; Gaps 20;

OY 20 SAASGSGQSTRYWDCKPSCAMPK--AAVSQPVYACDANFORLSDPNVSGCNGG---- 73
DB 211 SANAGTG---NYGTCCEMDIWEANNDAAAYTP-HPCIT-----MAOTRSGSCTR 258
OY 74 -----SAYSCADOTPMVAVNDNLAVGPAATSIAGSSSSWCCACALTTTSGPV 121
DB 259 DTGLCDADGCGDFSRMGQDT-----FLGKGLVDTSKPTVYVQFIT-NDGTS 306
OY 122 AG-----KTMVVOSTGTGDLGNOFDIAMPG-GVGIF--NCCSSQ---FGSLP-GAQT 169
DB 307 AGLTLEIRRLRYVN---GKVIQNS-SVKIPGIDPNASTITDNCSSOKTAFGDTNFAOH 361
OY 170 GGISRHQDQ-----DSFPAPLKPQCGMRFDFQ-----NADNPTFFFOQVOC- 211
DB 362 GGLKQVGEALRTGMVLALSIWIDYANML-----WIDSNTPTKKDPSTPEVAGTCA 413
OY 212 -----PAEIVASGCKRNDSSFPVTPPSGGNGT-----GTPSTAP 250
DB 414 TMSGVAQIEAQS-----PNAVYVESNFKFGDLNTTYTGTVSSSSVSSSSSTSSSH 467
OY 251 GSGQSTRYWDCKPSCAMPKGAASQPVYACDANFORLSDPNVSGCNGSAYSCADOTP 83
DB 468 SSSSTPTPTPTGTVTVFPMQCGCGIGTGTTCACSPYTCVHLNPNYVQC 515

RESULT 12
GUX1_PENJA STANDARD; PRT; 537 AA.
ID GUX1_PENJA
AC 006886;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DE 15-JUL-1999 (Rel. 38, Last annotation update)
DE Exoglucanase I precursor (EC 3.2.1.91) (Exocellobiohydrolase I)
GN CBH1.
OS Penicillium janthinellum (Penicillium vitale).
OC Eukaryotes; Fungi; Ascomycota; Peizizymycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic trichocomaceae; Penicillium.
OX NCBI_Taxid=5079;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CA1;
RX MEDLINE=93178976; Pubmed=8440481;
RA Koch A., Weigel C.T.O., Schulz G.;
RT Cloning, sequencing, and heterologous expression of a cellulase-
RT encoding cDNA (cbh1) from Penicillium janthinellum.";
RL Gene 124:57-65(1993).
CC -1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
CC in cellulose and cellobiose, releasing cellobiose from the non-
CC reducing ends of the chains.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY C (FAMILY 7 OF GLYCOSYL
CC HYDROLASES).
CC -1- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X59054; CAA1780.1; -;
DR PIR: J00150; J00150.
DR HSSP: P00725; 8CEL.
DR InterPro: IPR000254; CBD_fungal.
DR InterPro: IPR001722; GH_7.
DR Pfam: PF00734; CBM_1; 1.
DR Pfam: PF00840; Glyco_hydro_7; 1.
DR ProDom: PD001821; CBD_fungal; 1.
DR ProDom: PD186135; GH_7; 1.
DR SMART: SM00236; fCBD; 1.
DR PROSITE: PS00562; CBD_FUNGAL; 1.
KM Cellulose degradation; Hydrolyase; glycoprotein; signal.
FT SIGNAL 1
FT CHAIN 19 537
FT DOMAIN 19 453
FT DOMAIN 454 477
FT DOMAIN 478 537
FT ACT_SITE 235 235
FT ACT_SITE 240 240
FT CARBOHYD 136 136
FT CARBOHYD 414 414
FT CARBOHYD 456 456
FT DISULFID 509 526
FT DISULFID 520 536
SQ SEQUENCE 537 AA; 56842 MW; A6B9C6EB73F17FE4 CRC64;

Query Match 11.0%; Score 183; DB 1; Length 537;
Best Local Similarity 23.4%; Pred. No. 1,7e-06;
Matches 79; Conservative 30; Mismatches 99; Indels 130; Gaps 14;

OY 24 GSGQSTRYWDCKPSCAMPKGAASQPVYACDANFORLSDPNVSGCNGSAYSCADOTP 83
DB 268 GTVSTDRYGGTCDDP-----GCDENPYRMGVNTNFY-----GPGETIDRKSP 308

```

OY 84 WAW-----NDMLATGFAA-----TSINGGS-ESSWCCACATLTF 116
DB 309 FTVVTFITNDSTSTSEIKREYVQGGKVIKNPOSTIVGSGNSITDSWCNA----- 362
OY 117 TSGPAVAKTMVQSTGTGDLGNSQFDI--AMPGGVGIPNGC----- 157
DB 363 -----QKSAFDD--TNEFSKHGGMAGMGAGLADGAVIIVMSLMDHADSMLML 407
OY 158 -----SSQFGILPAGVAGYGISRDGCDSEFPAPLKPCCQMFDFQADNFTFTFOYQCP 212
DB 408 DSTYPTNATSTTPGAK-----RGTCDISRRP-----NTVESTY-----P 441
OY 213 AATVARSQGRNDSEFPVFTPPSGNGGCTGPTSTAPGSGQ-----TSPGGGS 261
DB 442 NAYVIYSNIRKGTPLNS--TFTGCTTSSSTTTTTSKSTSTSSSKTTTYYTTTTSSSGS 499
OY 262 GCTSOAKMACGCGIGFSGCTTGVSGTTCOKLNDYSSQL 299
DB 500 GTGARNDMACGGMGWTGPTTVCVSPYCTCKNDMYSOCL 537

RESULT 13
GUX1_NEUCR STANDARD; PRT; 516 AA.
AC P38676:
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DE 15-JUL-1999 (Rel. 38, Last annotation update)
DE Exoglucanase 1 precursor (EC 3.2.1.91) (Exocellulohydrolase 1)
GN CBH-1.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=74-OR23-1A;
RX MEDLINE=95369725; PubMed=7642129;
RA Taleb F., Radford A.;
RT "The cellulase complex of Neurospora crassa: cbh-1 cloning,
RT sequencing and homologies."
RL Gene 161:137-138(1995).
CC -1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
CC in cellulose and cellobiose, releasing cellobiose from the non-
CC reducing ends of the chains.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY C (FAMILY 7 OF GLYCOSYL
CC HYDROLASES).
CC -1- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X77778; CAA54815.1; -.
CC PIR: S42093; S42093.
CC HSSP: P00725; 8CEL.
CC InterPro: IPR000254; CBD_fungal.
CC InterPro: IPR001722; GH_7.
CC Pfam: PF00734; CBM_1; 1.
CC Pfam: PF00840; Glyco_hydro_7; 1.
CC ProDom: PD001821; CBD_fungal; 1.
CC ProDom: PD186135; GH_7; 1.
CC SMART: SM00236; fcbd; 1.
CC Cellulose degradation; Hydrolase; Glycosidase; Glycoprotein; Signal.
CC STGNL 1 17
CC CHAIN 18 516 EXOGLUCANASE 1.

```

```

FT DOMAIN 18 445 CATALYTIC.
FT DOMAIN 446 480 LINKER.
FT ACT_SITE 481 516 CELLULOSE-BINDING (BY SIMILARITY).
FT ACT_SITE 223 223 NUCLEOPHILE (BY SIMILARITY).
FT ACT_SITE 228 228 PROTON DONOR (BY SIMILARITY).
FT DISULFID 488 505 BY SIMILARITY.
FT DISULFID 499 515 BY SIMILARITY.
SQ SEQUENCE 516 AA; 54471 MW; 38E598406EA81900 CRC64;

Query Match 10.9%; Score 181; DB 1; Length 516;
Best Local Similarity 24.5%; Pred. No. 2.3e-06;
Matches 81; Conservative 43; Mismatches 129; Indels 78; Gaps 16;

OY 22 ASGSGSTRKYMDCCKPSCAW-PGKAASQPVYACDANFQRLSDFNVOGCGNGSAYS--- 77
DB 211 ANGID--HGSCCBMDIWEANKYSTATPHPCRTIEQHMCB-----GDSGCGTSSDDR 262
OY 78 ---C---ADQTPMAVNDNLAYGFAATSIAGSSESWCACALFTSGPVAG-KTMV 128
DB 263 YGVLCADAGCDNSRYRMGNTTFYEGEKTIV---DTSKFTVTFIDISADLAIEIAFY 319
OY 129 OSTSTGDLGNSQFDIAMPGGVGIFNGCSSQ---FGILPG-AQYGISRDQ----- 178
DB 320 QN---GKVTENSQSNVDGYSNSITQSFCKSQTAFGIDIDFKKKGSLQMGKALQAMV 376
OY 179 -----DSFPAPLKPCCQMFDFQADNFTFTFOYQCPAETVARSQC 221
DB 377 LVMSITMDHANMLMWDSTYTPVPKVEGA-----YRSGPPTS---GVPAEVDANAPN 425
OY 222 KRNDSSFPV---FTPPSGGNGT-----GTPT-STAPSSGQSTPSSGGGCTSQKV 268
DB 426 SKVAFSNIKRFGLGISPFSGSSGTPSPNPSSASPTSPAKPSSTASNPSTGAHW 485
OY 269 AOCGGIGFSGCTTGVSGTTCOKLNDYSSQL 299
DB 486 AOCGGIGFSGPTTCEPPTCANDHDITSCV 516

RESULT 14
CELL_AGABI STANDARD; PRT; 320 AA.
ID CELL_AGABI
AC 000023:
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cellulose-growth-specific protein precursor.
GN CELL.
OS Agaricus bisporus (Common mushroom).
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OC Agaricales; Agaricaceae; Agaricus.
OX NCBI_TaxID=5341;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D649;
RX MEDLINE=93012985; PubMed=1398098;
RA Raguz S., Yaguee E., Wood D.A., Thurston C.F.;
RT "Isolation and characterization of a cellulose-growth-specific gene
RT from Agaricus bisporus."
RL Gene 119:183-190(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=D649;
RX MEDLINE=94237428; PubMed=8181702;
RA Armesilla A.L., Thurston C.F., Yaguee E.;
RT "CELL: a novel cellulose binding protein secreted by Agaricus
RT bisporus during growth on crystalline cellulose."
RL FEMS Microbiol. Lett. 116:293-299(1994).
CC -1- FUNCTION: PROBABLE GLYCOSYL HYDROLASE ACTIVE ON CELLULOSE.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).
CC -1- SIMILARITY: BELONGS TO FAMILY 61 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb-sdb.ch).

DR EMBL: M86356; AAS3434.1; -
 DR HSSP: P00725; 2CBH.
 DR InterPro: IPR000254; CBD_fungal.
 DR InterPro: IPR005103; Glyco_hydro_61.
 DR Pfam: PF00734; CBM_1.1.
 DR Pfam: PF03443; Glyco_hydro_61.1.
 DR SMART: SM00236; fCBD.1.
 DR PROSITE: PS00562; CBD_FUNGAL; 1.
 DR Cellulose degradation: Hydrolase; Glycosidase; Glycoprotein; signal.
 FT SIGNAL 1 29
 FT CHAIN 30 320
 FT DOMAIN 262 285
 FT DOMAIN 286 320
 FT DOMAIN 292 309
 FT DISULFID 303 319
 FT CARBOHYD 163 163
 SO SEQUENCE 320 AA: 33754 MW: 602CB080895CA2B CRC64:

Query Match 10.4%; Score 172; DB 1; Length 320;
 Best Local Similarity 24.8%; Pred. No. 6.6e-06;
 Matches 85; Conservative 36; Mismatches 132; Indels 90; Gaps 15;

QY 10 TLAAALPLVANAASGSGSTRY-WDCC-----KPSGAMPKRAVSQVYACDANFQRLSD 63
 Db 15 TFSLALGLFAAKVAOHAGVIGVSMGTWYEGWHPYNTFPGQTSIRPW---ATFDPIMD 70
 QY 64 FNVOS-CGNGSAYSCADQTPMAVNDLAYSFAATSIAGSSSSSCCACTALFTSSGYA 122
 Db 71 ATATVYGCNN-----DGNP-GRNQLTATVAAAGTATTAIVNQPMPHYGPTTYLKGCP 122
 QY 123 KRTWVOSTST-----GDDLG-----SNQDPIAMPGGVGIFNGCS 158
 Db 123 GSSCDGVNTNLSKWFKEIDAGLLSTGVKGWVGSKMIDNNMSTTTP- 171
 QY 159 SQFGPLPAGVAGISRRQCSFAPPLKPGCWMPFQMDNPTFTQVQCP--AEIV 216
 Db 172 ---STVSGAVMIRPETHALSHLPAQIYPECA-----QLTTGCGNRAPSSSELY 218
 QY 217 ARSGCKRND-----DSSEPVFTTP---SGNGG--TGPTSTAPSGQTS 256
 Db 219 SFGGYSNDPGLVNLVTQELAMDTTYIIVCPPLLYGSGNGSFTTPTTPTTSP 278
 QY 257 PGGSGCTSOAKMAOCGIGFSGCTTCVSGTTQCKLNDYSSQL 299
 Db 279 PTSTPG-TIPYGGCGGIGTGTGCVAPYQCKVIINDYSSQL 320

RESULT 15
 GUNA_TIRE STANDARD; PRT; 344 AA.
 AC 014405;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DE Endoglucanase IV precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase IV)
 GN (cellulase IV) (EGIV).
 GN EGL4.
 OS Trichoderma reesei (Hypocrea jecorina).
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Hypocreales; Hypocreaceae; Hypocrea.
 OX NCBI_TaxID=51453;
 RN [1]
 RP SEQUENCE FROM N.A., ACTIVITY, AND INDUCTION.
 RC STRAIN-QM9414 / Rut C-30;
 RX MEDLINE=98036137; PubMed=9370370;

RA Saloheimo M., Nakari-Setälä T., Tenkanen M., Penttilä M.;
 RT "cDNA cloning of a Trichoderma reesei cellulase and demonstration of
 RT endoglucanase activity by expression in yeast."
 RL Eur. J. Biochem. 249:584-591(1997).
 CC -1- FUNCTION: The biological conversion of cellulose to glucose
 CC generally requires three types of hydrolytic enzymes:
 CC (1) Endoglucanases which cut internal beta-1,4-glycosidic bonds;
 CC (2) Exocellulohydrolases that cut the disaccharide polymer chain;
 CC (3) Beta-1,4-glucosidases which hydrolyze the cellobiose and other
 CC short cello-oligosaccharides to glucose.
 CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
 CC linkages in cellulose.
 CC -1- SUBCELLULAR LOCATION: secreted.
 CC -1- INDUCTION: By cellulose, cellobiose, lactose and sophorose.
 CC -1- PTM: May also be O-glycosylated.
 CC -1- SIMILARITY: BELONGS TO FAMILY 61 OF GLYCOSYL HYDROLASES.
 CC -1- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN
 CC (CBD).

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb-sdb.ch).

DR EMBL: Y11113; CAA71999.1; -
 DR HSSP: P00725; 2CBH.
 DR InterPro: IPR000254; CBD_fungal.
 DR InterPro: IPR005103; Glyco_hydro_61.
 DR Pfam: PF00734; CBM_1.1.
 DR Pfam: PF03443; Glyco_hydro_61.1.
 DR ProDom: PD001821; CBD_fungal; 1.
 DR SMART: SM00236; fCBD.1.
 DR PROSITE: PS00562; CBD_FUNGAL; 1.
 DR Cellulose degradation: Hydrolase; Glycosidase; Glycoprotein; signal.
 FT SIGNAL 1 21
 FT CHAIN 22 344
 FT DOMAIN 22 256
 FT DOMAIN 257 307
 FT DOMAIN 308 344
 FT DISULFID 315 332
 FT DISULFID 326 342
 FT CARBOHYD 80 80
 FT CARBOHYD 158 158
 SO SEQUENCE 344 AA: 35510 MW: 7BFLC4AB705350C CRC64;

Query Match 9.6%; Score 159; DB 1; Length 344;
 Best Local Similarity 24.1%; Pred. No. 6.3e-05;
 Matches 71; Conservative 31; Mismatches 102; Indels 90; Gaps 14;

QY 41 WPKRAVSQVYACDAN-----FQRLSDPFWVSGCNGSAYS---CADQTPMAVN 87
 Db 105 WPHGPIVDYLANCGDEFVDTKTLFEFKIDGVLLSGDRGTASDVLLSNNTWYK 164
 QY 88 --DNLAVGFATSIAGSSSSSCCACTALFTSGVAKTWNVOSTSTGGLGSGNQ----- 141
 Db 165 IPDNLAPG-----NYVL-----RHETIHLHSAAGANGANYPQC 198
 QY 142 FDIAMPGGVGIFNGCSQFGLPGAQ-YGGISRHQCDSPFAPLK---PGQWRFDMFQ 197
 Db 199 FNIIVSGG-----SLQPSGVLTDLVHMTDRGVLLINITYSPNTIIPG----- 242
 QY 198 MADNPTFTFQVQCPAEIVARSGCKRNDSSFPVTPPSGNGGTGPTSTA----- 249
 Db 243 ---PTVV---SGLPTSVAGSSAATATASA---TVPGGSGSPTRTTTARTTQASSR 291
 QY 250 ---PGSGOTSPGGSGCTSOAKMAOCGIGFSGCTTCVSGTTQCKLNDYSSQL 299
 Db 292 PSTTPPATTSAPAGP---TQTLVGGCGGSGTSGPLRCAPATCTSLNBYAQC 343

Fri Feb 28 12:24:07 2003

us-10-007-521-12.rsp

Page 12

Search completed: February 24, 2003, 11:51:53
Job time : 16 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 24, 2003, 11:49:14 ; Search time 33 Seconds

(without alignments)
1866.912 Million cell updates/sec

Title: US-10-007-521-12

Perfect score: 1659

Sequence: 1 MRSTPVLRTTLAALPLVAS.....TTCVSGTTCCKLNDYISQCL 299

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: SP_ARCHAEA:*
2: SP_BACTERIA:*
3: SP_FUNGI:*
4: SP_HUMAN:*
5: SP_INVERTEBRATE:*
6: SP_MAMMAL:*
7: SP_MHC:*
8: SP_ORGANELLE:*
9: SP_PHAGE:*
10: SP_PLANT:*
11: SP RODENT:*
12: SP_VIRUS:*
13: SP_VERTEBRATE:*
14: SP_UNCLASSIFIED:*
15: SP_VIRUS:*
16: SP_BACTERIAP:*
17: SP_ARCHAEP:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1170	70.5	305	3	093782 humicola gr
2	724	43.6	219	14	09JH92 unclassified
3	724	43.6	220	14	09JH84 unclassified
4	709	42.7	219	14	09JH91 unclassified
5	706.5	42.6	220	14	09JH83 unclassified
6	705	42.4	217	14	09JH95 unclassified
7	703	42.4	219	14	09JH86 unclassified
8	700	42.2	219	14	09JH87 unclassified
9	698.5	42.1	227	3	093783 humicola gr
10	698	42.1	219	14	09JH90 unclassified
11	697.5	42.0	221	14	09JH94 unclassified
12	697	42.0	219	14	09JH89 unclassified
13	693	41.8	220	14	09JH88 unclassified
14	689.5	41.6	220	14	09JH96 unclassified
15	689.5	41.6	220	14	09JH93 unclassified
16	680.5	41.0	218	14	09JH85 unclassified

17	631	38.0	410	3	09P868 pitomyces e
18	585.5	35.3	271	3	09UVP3 altermaria
19	543.5	32.8	242	5	097401 phaedon coc
20	360.5	21.7	112	14	09JH99 unclassified
21	348.5	21.0	112	14	09JH40 unclassified
22	347.5	20.9	112	14	09JH98 unclassified
23	338.5	20.4	112	14	09JH41 unclassified
24	338.5	20.4	112	14	09JH42 unclassified
25	330	19.9	111	14	09JH97 unclassified
26	247	14.9	353	3	09HE18 penicillium
27	215	13.0	514	3	093832 trichoderma
28	200	12.1	505	3	09P883 trichoderma
29	191	11.5	525	3	012621 humicola gr
30	188.5	11.4	302	3	099034 trichoderma
31	183	11.0	523	3	09Y723 trichoderma
32	181.5	10.9	529	3	08W234 trichoderma
33	176	10.6	283	3	096UV7 penicillium
34	169.5	10.2	536	3	09UVS8 penicillium
35	163.5	9.9	536	3	09Y895 aspergillus
36	163	9.8	412	3	074169 volvariella
37	161	9.7	397	3	08TFB0 aspergillus
38	161	9.7	504	3	001763 trichoderma
39	161	9.7	516	3	096VU3 phanerochaete
40	160	9.6	517	3	09Y722 lentinula e
41	156	9.4	856	3	074170 aspergillus
42	152.5	9.2	290	3	09HE20 phanerochaete
43	150.5	9.1	526	3	09Y724 trichoderma
44	150.5	9.1	860	3	002290 neocallimastix
45	150	9.0	479	3	09UUS4 colleotric

ALIGNMENTS

RESULT 1

093782 ID: 093782 PRELIMINARY: PRT: 305 AA.
AC 093782;
DT 01-MAY-1999 (TREMUREL. 10, Created)
DT 01-MAY-1999 (TREMUREL. 10, Last sequence update)
DT 01-JUN-2002 (TREMUREL. 21, Last annotation update)
DE Endoglucanase.
GN EGU3.
OS Humicola grisea var. thermoldea.
OC Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Humicola.
OX NCBI_TaxID=5528;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IFO9854;
RX MEDLINE=99144540; PubMed=990729;
RA Takashima S., Ikura H., Nakamura A., Hidaka M., Masaki H., Uozumi T.;
RT "Comparison of gene structures and enzymatic properties between two
endoglucanases from Humicola grisea";
RL J. Biotechnol. 67:85-97(1999).
DR EMBL; AB003107; BAA74956.1; -.
DR HSSP; P43316; ZENG.
DR InterPro; IPR000254; CBD_fungal.
DR Pfam; PF00734; CBM_1; 1.
DR Pfam; PF02015; Glyco_hydro_45; 1.
DR ProDom; PD001821; CBD_fungal; 1.
DR SMART; SM00236; fcbd; 1.
DR PROSITE; PS00562; CBD_FUNGAL; 1.
DR PROSITE; PS01140; GLYCOSYL_HYDROL_F45; 1.
SQ SEQUENCE 305 AA: 32174 MW: 28C979DEDCD771D CRC64;

Query Match 70.5%; Score 1170; DB 3; Length 305;

Best Local Similarity 67.8%; Pred. No. 4e-85;

Matches 208; Conservative 44; Mismatches 45; Indels 10; Gaps 4;

QY 1 MRSTPVLRTTLAALPLVASAGSGSTRVWDCKPSCAMPKRAVSPVYACDANQR 60
DB 1 MRSSPLPSDVVALPVLALAA--DGKSTRVWDCKPSCGMAKRAVPVQPVFSCNANQR 58

Oy	61	LSDPNVSGCN-GGSAYSCADQFMVAANDMLAVGFAATSTAGSSESSGCACALFTTSG	119
Dd	59	LTPDASGCEGPVAAISCADQFMVAANNDDPFGAFATSTAGSNEAWCCACTELFTTSG	118
Oy	120	PVAKTIWVOYSTGTGGDLGSNQDIAMPGGGVGIPIGCCSSQFGLPGEAQYGGIISRDCD	179
Dd	119	PVARKKMYVOSTSTGGGLGSHNEFLNLPGGGVGFIPDCCTQFQGLPGQAQRGGISRNEDC	178
Oy	180	SFPALKPGCOMRFDMFNADNPFTPTFOYOQCPEAIYARSGCKRNDSPSP-FETPPSG	238
Dd	179	RFPALKPGCYWRDWMKNADNPFSFRJOQCRAELVARIGCRNDNGNFPANQIPBSST	238
Oy	239	NGGTGTPTSTAIPSGGTS-----PGGSGSCTISOKWAQCGGIGBSGTCVCSTCOCLN	292
Dd	239	SSPVGQPTSTSTSTSTSSBPVQPTTPSGCTAERMAOCCGNGMSGCTTCVAGSTCTKIN	298
Oy	293	DYVSQCL	299
Dd	299	DWIHQCL	305

RESULT 2	09JH92	PRELIMINARY;	PRT;	219 AA.
ID	09JH92			
AC	09JH92;			
DT	01-OCT-2000 (TREMBLrel. 15, Created)			
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)			
DT	01-MAR-2002 (TREMBLrel. 20, Last annotation update)			
DE	Family 45 cellulase homologue.			
OS	Unclassified eukaryotes.			
OC	Eukaryota.			
OX	NCBI_TaxID=42452;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Ohkoko K., Ohkuma M., Moriya S., Kudo T.;			
RT	"Diverse genes of family 45 cellulase homologues of the symbiotic protists in the hindgut of termite <i>Reticulitermes speratus</i> ."			
RL	Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AB045171; BAA98041.1; -.			
DR	HSSP; P43316; 2ENG.			
DR	InterPro; IPR000334; GH_45.			
DR	Pfam; PF02015; Glyco_hydro_45; 1.			
DR	PROSITE; PS01140; GLYCOSYL_HYDROL_F45; UNKNOWN_1.			
SQ	SEQUENCE 219 AA; 23001 MW; 5f2EB81A6DE926CE CRC64;			
Query Match	43.6%; Score 724; DB 14; Length 219;			
Best Local Similarity	60.3%; Pred. No. 6, 2e-50;			
Matches 132; Conservative	26; Mismatches 53; Indels 8; Gaps			
QY	11 LAALPLVSAASAGSGSTRWYDCKPSCAMPKGAVSQPVYAC--DANFORLSDFNNVOS 68			
Db	2 LVFIALILSYFGSDGRTRWYDCKKASCAMKEKKAALVPTDVGKDTTRVAASNDTVKS 61			
QY	69 GCGNSAATSCADQDTWYAVNDNLAVFAATSTNGSESSSMCCACVYALFTSSGVAGKTMV 128			
Db	62 ACDGEGEWCYDQAAWAAVNDVAVFAAACCAG--ESGACNCMEYELTFTSSGPVNGKRMV 120			
QY	129 QSTSTGDLGNSQNPFIAMPGGVGIFNCSSQFGGLP---GAOYGLGISRPQCDSEFPAP 184			
Db	121 QVTNNGDLGNSQNPFLAIPGGVGIGYNGCTQ--SGAPADGMSRRYGGVSRSECSQLPSG 179			
QY	185 LKPGCAQRFDMFQNDANPTFTFOQVQCPAEIYASGCKR 223			
Db	180 LQAGCQMRFDWFQNDANPISINQYTCPELPYAKTNCKR 218			
RESULT 3				
ID	09JH84	PRELIMINARY;	PRT;	220 AA.
AC	09JH84;			
DT	01-OCT-2000 (TREMBLrel. 15, Created)			
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)			

DT 01-MAR-2002 (TIREB1refl. 20, last annotation update)
DE Family 45 cellulase homologue.
OS unclassified eukaryotes.
OC Eukaryota.
OX NCBI_TaxID=42452;
RN [1]
RP SEQUENCE FROM N.A.
RA Ohtoko K., Ohkuma M., Moriya S., Kudo T.;
RT "Diverse genes of family 45 cellulase homologues of the symbiotic
RT protists in the hindgut of termite *Reticulitermes speratus*.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB045179; BAA98049.1; -;
DR HSSP; P43316; ZENG.
DR InterPro; IPR000334; GH_45.
DR Pfam; PF02015; Glyco_hydro_45; 1.
DR PROSITE; PS01140; GLYCOSYL_HYDROL_F45; UNKNOWN_1.
SQ SEQUENCE 220 AA; 23274 MW; 26AF557512E061 CRC64;

[illegible]

RESULT 4			
09JH91			
ID	09JH91	PRELIMINARY;	PRT; 219 AA.
AC	09JH91.		
DT	01-OCT-2000 (TREMBLrel. 15, Created)		
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)		
DT	01-MAR-2002 (TREMBLrel. 20, Last annotation update)		
DE	Family 45 cellulase homologue.		
OS	unclassified eukaryotes.		
OC	Eukaryota.		
OX	NCBI_TaxID=42452;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Ohnoko K., Ohnuma M., Moriya S., Kudo T.;		
RT	"Diverse genes of family 45 cellulase homologues of the symbiotic		
RT	protists in the hindgut of termite Reticulitermes speratus.";		
RL	Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.		
DR	EMBL: AB045172; BAA98042.1; -.		
DR	HSSP: P43316; 2ENG.		
DR	InterPro: IPR000334; GH_45.		
DR	Pfam: PF02015; Glyco_hydro_45.1.		
DR	PROSITE: PS01140; GLYCOSYL_HYDROL_F45; UNKNOWN_1.		
SO	SEQUENCE 219 AA; 23126 MW; A112EF3FCAB041C CRC64;		

[illegible]

[illegible]

RESULT 5	
Q9JH83	
ID Q9JH83	PRELIMINARY;
Q9JH83	PRT; 220 AA.

DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE Family 45 cellulase homologue.
OS unclassified eukaryotes.
OC Eukaryota.
NCBI_TaxID=42452;
RN
RP
RF
RA Ontoko K., Ohkuma M., Moriya S., Kudo T.;
RT "Diverse genes of family 45 cellulase homologues of the symbioida
RT protists in the hindgut of termites Reticulitermes speratus."
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB045167; BAA98037.1; -
DR EMBL: AB045166; BAA98036.1; -
DR HSSP: PA3316; 2ENG.
DR InterPro: IPR000334; GH_45.
DR Pfam: PF02015; Glyco_hydro_45.1.
DR PROSITE: PS01140; GLYCOSYL_HYDROL_F45; UNKNOWN_1.
SQ
SEQUENCE 220 AA: 23108 MW: 97738DB831BCFA5F3 CRC64;

Query Match	42.68;	Score 706.5;	DB 14;	Length 220
Best Local Similarity	57.98;	Pred. No. 1.5e-48;		

Matches 128; Conservative 32; Mismatches 52; Indels 9; Gaps 5;

Qy 1.1 LAALPLVASAASG-SGOSTRYNDCCKPSCAWPGKAASQPVAC--DANFORLSDFNVQ 67

Db 2 LAFVPSLLASVLEFGDSCKTTRYNDCCKSGSGMEAKADVSKPIOTCAKDGTTIRVASNDTVK 61

Qy 68 SGCGNGSAYSCADQTFPWAUNDNLAYGPAATSLAGGSESSWCCACALYTFMSGPVAGKTMV 127

Db 62 SGCGDGGDGYCYDQTFPMGVNDSTALGFAAAALISGG-EKAACCNCELYLTFMSGPVNGKMT 120

QY 128 VQSTGTGDLGSNQFDIAMPGCGVGIRNGCSSQFGGLP---GAQYGGISSRDQCDSPFA 183

Db 121 VQVNTGTGDLGSNQFDIAPGCGVGIRNGCTAO-SGAPADGWSRRYGVSSRSCQLPS 179

QY	184	PLKPGCQWRFDWFQADNP	PTFTFQOVCPAEIVARSGCKRN	224
		:	: : : : :	
Db	180	GIAGCGWRFDWFQADNP	SPSMENFVYSCPSFLAKTNCBRN	220

RESULT 6	
Q9JH95	
ID Q9JH95	PRELIMINARY;
	PRT; 217 AA.

DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE Family 45 cellulase homologue.
OS unclassified eukaryotes

NCBI_TaxID=42452;

RN [1]

RA Ohtoko K., Ohkuma M., Moriya S., Kudo T.

RT prolists in the hindgut of termite *Reticulitermes speratus*.";

RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases
 DR EMBL: AB045169; BAA98038.1; -
 DR HSSP: P43316; 2ENG.
 DR InterPro: IPR000334; GH_45.
 DR Pfam: PF02015; Glyco_hydro_45; 1.
 DR PROSITE: PS01140; GLYCOSYL_HYDROL_F45; UNKNOWN_1
 SO SEQUENCE 217 AA; 22796MW; 660DD1336B3769DB CRC64;

Query Match	42.5%	Score 705	DB 14	Length 217
Best Local Similarity	58.7%	Pred. No. 2e-48		
Matches 125	Conservative 33	Mismatches 47	Indels 8	Gaps 4

QY 18 VASAASSGOSTRYWDCPCPSCAWPGKAAVSQPYAC--DANFRLSDFNVOGCNGSSA 75
::|||::||||||| || | |||: | : | :|||:|

Oy 76 YSCADQTPMAVNDNLAVGAATSTAGGSESSWCACALFTSPVA GKTMVOYSTGG 135
:
Db 67 FMCIDQTPMGVNDSTALGFALAAISGG-EKRAACNCIEFLPTSPVAGKKMTYOVYTNGG 125

```
Qy      136 DLGSGQFDLAMPGGVGIFNCGSSQFGLP-----GAQYGGISSRQCDSFAAPLKPCCQM 191
         |||||:::|||||::|||::: | ::|||::: : | ::|||
Db      126 DLGSGQFDLAIPGGVGIYNCTAQ--SCAPADNGSRYGVSRSRCQSQTLSGLQGACQM 184
```

```
QY 192 RFDWQONADNPFTFTQQVQCPAEIIVARSCKRN 224
      |||||: | | | : : : | : ||
Db 185 RFDWQONADNPMSNFNVVSCPSIELIAKTICRRN 217
```

RESULT 7
Q9JH86
ID Q9JH86 PRELIMINARY; PRT; 219 AA

DR 01-OCT-2000 (TREMBrel. 15, Created)
 DT 01-OCT-2000 (TREMBrel. 15, Last sequence update)
 DT 01-MAR-2002 (TREMBrel. 20, Last annotation update)
 DE Family 45 cellulase homologue.
 OS unclassified eukaryotes.

OC Eukaryota.

RN [1]
RP SEQUENCE FROM N.A.
RA Ohtoko K., Ohkuma M., Moriya S., Kudo T.;

RT protists in the hindgut of termite *Reticulitermes speratus*.";
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB045177; BAA98047.1; -

DR InterPro: IPR000334; GH_45.
DR Pfam: PF02015; Glyco_hydro_45; 1.
DR PROSITE: PS01140; GLYCOSYL_HYDROL_F45; UNKNOWN 1.

Query Match	Score	DB	Length
Best Local Similarity	42.48	703	219
	58.18	2	9e-48

```

qy      15 LPLVASAAGSGGQSTRYWDCCAPSCAWMGKAAYSQPVYAC--DANFQRLSPFNWQSGCG 72
      | : : : : : | | | | | : : : : : | : : : : : |
Db      6 LTFVGLSLAEGKTTTRYWDCCGSGCGWEKKANNDKPIDCAKDGTTTRYVASNDTVKSGCDG 65

```

[illegible]

Db 125 TGGDLGSNQFDLAI PGGCVGIYNGCTAQ-SCAPSDGWGSS

```
RESULT 8
09JH87 PRELIMINARY; PRT; 219 AA.
ID 09JH87:
AC 01-OCT-2000 (T-EMBLrel. 15, Created)
DT 01-OCT-2000 (T-EMBLrel. 15, Last sequence update)
DT 01-MAR-2002 (T-EMBLrel. 20, Last annotation update)
DE Family 45 cellulase homologue.
OS unclassified eukaryotes.
OC Eukaryota.
NCBI_TaxID=42452;
RN [1]
RX SEQUENCE FROM N.A.
RA Ohtoko K., Ohkuma M., Moriya S., Kudo T.;
RT "Diverse genes of family 45 cellulase homologues of the symbiotic
RT protists in the hindgut of termite Reticulitermes speratus.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB045176; BAA98046.1; -.
DR HSSP; P43316; ZENG.
DR InterPro; IPR000334; GH_45.
DR Pfam; PF02015; Glyco_hydro_45; 1.
DR PROSITE; PS01140; GLYCOSYL_HYDROL_F45; UNKNOWN_1.
SQ SEQUENCE 219 AA; 23134 MW; 4BDEF4EC9ACC772D CRC64;

Query Match 42.2%; Score 700; DB 14; Length 219;
Best Local Similarity 57.7%; Pred. No. 4,9e-48;
Matches 124; Conservative 34; Mismatches 49; Indels 8; Gaps 4;

QY 15 LPLVASAASGSGSTRYWDCKPSCAMPKAAVSQPYAC--DANFQRLSDFNVOGSCNG 72
ID 09JH90: PRELIMINARY; PRT; 219 AA.
AC 09JH90:
DT 01-OCT-2000 (T-EMBLrel. 15, Created)
DT 01-OCT-2000 (T-EMBLrel. 15, Last sequence update)
DT 01-MAR-2002 (T-EMBLrel. 20, Last annotation update)
DE Family 45 cellulase homologue.
OS unclassified eukaryotes.
OC Eukaryota.
NCBI_TaxID=42452;
RN [1]
RX SEQUENCE FROM N.A.
RA Ohtoko K., Ohkuma M., Moriya S., Kudo T.;
RT "Diverse genes of family 45 cellulase homologues of the symbiotic
RT protists in the hindgut of termite Reticulitermes speratus.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB045173; BAA98043.1; -.
DR HSSP; P43316; ZENG.
DR InterPro; IPR000334; GH_45.
DR Pfam; PF02015; Glyco_hydro_45; 1.
DR PROSITE; PS01140; GLYCOSYL_HYDROL_F45; UNKNOWN_1.
SQ SEQUENCE 219 AA; 23037 MW; 372E016415530A9A CRC64;

RESULT 9
09JH83 PRELIMINARY; PRT; 227 AA.
ID 09JH83:
AC 09JH83:
DT 01-MAY-1999 (T-EMBLrel. 10, Created)
DT 01-MAY-1999 (T-EMBLrel. 10, Last sequence update)
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE Endoglucanase.
GN EglA.
OS Humicola grisea var. thermolidea.
OC Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Humicola.
NCBI_TaxID=5528;
RN [1]
RX SEQUENCE FROM N.A.
RA STRAIN-IFO9854;
RC MEDLINE=99144540; Pubmed=9990729;
RT Takashima S., Iikura H., Nakamura A., Hidaka M., Masaki H., Uozumi T.;
RT "Comparison of gene structures and enzymatic properties between two
RT endoglucanases from Humicola grisea.";
RL J. Biotechnol. 67:85-97(1999).
DR EMBL; AB003108; BAA74957.1; -.
DR HSSP; P43316; ZENG.
DR InterPro; IPR000334; GH_45.
DR Pfam; PF02015; Glyco_hydro_45; 1.
DR PROSITE; PS01140; GLYCOSYL_HYDROL_F45; UNKNOWN_1.
SQ SEQUENCE 227 AA; 24240 MW; 873553E76F5C39E4 CRC64;
```

```
Query Match 42.1%; Score 698.5; DB 3; Length 227;
Best Local Similarity 52.7%; Pred. No. 6.7e-48;
Matches 137; Conservative 26; Mismatches 46; Indels 51; Gaps 8;

QY 9 TTTLAALPLVASAASGSGSTRYWDCKPSCAMPKAAVSQPYACANQRLSD-FNVQ 67
ID 09JH94: PRELIMINARY; PRT; 221 AA.
AC 09JH94:
DT 01-OCT-2000 (T-EMBLrel. 15, Created)
DT 01-OCT-2000 (T-EMBLrel. 15, Last sequence update)
DT 01-MAR-2002 (T-EMBLrel. 20, Last annotation update)
DE Family 45 cellulase homologue.
OS unclassified eukaryotes.
OC Eukaryota.
NCBI_TaxID=42452;
RN [1]
RX SEQUENCE FROM N.A.
RA Ohtoko K., Ohkuma M., Moriya S., Kudo T.;
RT "Diverse genes of family 45 cellulase homologues of the symbiotic
RT protists in the hindgut of termite Reticulitermes speratus.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB045173; BAA98043.1; -.
DR HSSP; P43316; ZENG.
DR InterPro; IPR000334; GH_45.
DR Pfam; PF02015; Glyco_hydro_45; 1.
DR PROSITE; PS01140; GLYCOSYL_HYDROL_F45; UNKNOWN_1.
SQ SEQUENCE 219 AA; 23037 MW; 372E016415530A9A CRC64;

Query Match 42.1%; Score 698; DB 14; Length 219;
Best Local Similarity 56.7%; Pred. No. 7.1e-48;
Matches 122; Conservative 36; Mismatches 49; Indels 8; Gaps 4;

QY 15 LPLVASAASGSGSTRYWDCKPSCAMPKAAVSQPYAC--DANFQRLSDFNVOGSCNG 72
ID 09JH94: PRELIMINARY; PRT; 221 AA.
AC 09JH94:
DT 01-OCT-2000 (T-EMBLrel. 15, Created)
DT 01-OCT-2000 (T-EMBLrel. 15, Last sequence update)
DT 01-MAR-2002 (T-EMBLrel. 20, Last annotation update)
DE Family 45 cellulase homologue.
OS unclassified eukaryotes.
OC Eukaryota.
NCBI_TaxID=42452;
RN [1]
RX SEQUENCE FROM N.A.
RA Ohtoko K., Ohkuma M., Moriya S., Kudo T.;
RT "Diverse genes of family 45 cellulase homologues of the symbiotic
RT protists in the hindgut of termite Reticulitermes speratus.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB045173; BAA98043.1; -.
DR HSSP; P43316; ZENG.
DR InterPro; IPR000334; GH_45.
DR Pfam; PF02015; Glyco_hydro_45; 1.
DR PROSITE; PS01140; GLYCOSYL_HYDROL_F45; UNKNOWN_1.
SQ SEQUENCE 219 AA; 23037 MW; 372E016415530A9A CRC64;

RESULT 11
09JH94 PRELIMINARY; PRT; 221 AA.
ID 09JH94:
AC 09JH94:
DT 01-OCT-2000 (T-EMBLrel. 15, Created)
DT 01-OCT-2000 (T-EMBLrel. 15, Last sequence update)
DT 01-MAR-2002 (T-EMBLrel. 20, Last annotation update)
DE Family 45 cellulase homologue.
OS unclassified eukaryotes.
OC Eukaryota.
NCBI_TaxID=42452;
RN [1]
RX SEQUENCE FROM N.A.
RA Ohtoko K., Ohkuma M., Moriya S., Kudo T.;
RT "Diverse genes of family 45 cellulase homologues of the symbiotic
RT protists in the hindgut of termite Reticulitermes speratus.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB045173; BAA98043.1; -.
DR HSSP; P43316; ZENG.
DR InterPro; IPR000334; GH_45.
DR Pfam; PF02015; Glyco_hydro_45; 1.
DR PROSITE; PS01140; GLYCOSYL_HYDROL_F45; UNKNOWN_1.
SQ SEQUENCE 219 AA; 23037 MW; 372E016415530A9A CRC64;
```

```

AC 09JH94;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DE 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Family 45 cellulase homologue.
OS unclassified eukaryotes.
OC Eukaryota.
OX NCBI_TaxID=42452;
RN [1]
RP SEQUENCE FROM N.A.
RA Ohtoko K., Ohkuma M., Moriya S., Kudo T.;
RT "Diverse genes of family 45 cellulase homologues of the symbiotic
RL protists in the hindgut of termite Reticulitermes speratus.";
EMBL: AB045169; BAA98039.1; -
DR HSSP: P43316; 2ENG.
DR InterPro: IPR000334; GH_45.
DR Pfam: PF02015; GLYCOHYDROL_F45; UNKNOWN_1.
DR PROSITE: PS01140; GLYCOSYL_HYDROL_45; UNKNOWN_1.
SQ SEQUENCE 221 AA; 23030 MW; 8ABACEB0A8C46372 CRC64;

Query Match
Best Local Similarity 42.0%; Score 697.5; DB 14; Length 221;
Matches 124; Conservative 37; Mismatches 48; Indels 9; Gaps 5;

OY 14 ALPLVASASG-SGSTRYWDCKPSCAMPKAAVSQPVYAC--DANFORLSDFNVSQSG 70
DB 5 AYVLLALNLFSGSGKTRTYWDCKSGCWBAKADVSKPIDTCAKDDGTTTVAASNDTVKSGC 64
OY 71 NGASVASCADQTPMAVNDLAVGPAATSIAGSESSWCACALFTTSGPVAGKTVVOS 130
DB 65 DCGDGFMCYDQTPMAVNDLSIGFAAAVSGG-EKAAQCCELTFTSGPVNGKKTIVY 123
OY 111 TSTGDLGSGNPDIAIPGGGVGIFNGCSSQFGGLP---GAOYGISRRDQCSFPAPLK 186
DB 124 TMTGDLGSGNPDIAIPGGGVGIVNGCTAQ--SGAPADGMSRGVGSRSRSCQLPSGLQ 182
OY 187 PGCGQWRFDFWONADNPTFTFOQVOCPAETVARSCKRN 224
DB 183 AGCGWRFDFWONADNPTFTFNVSQPSLISKTNCRN 220

RESULT 12
OY 09JH89 PRELIMINARY; PRT; 219 AA.
AC 09JH89;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DE 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Family 45 cellulase homologue.
OS unclassified eukaryotes.
OC Eukaryota.
OX NCBI_TaxID=42452;
RN [1]
RP SEQUENCE FROM N.A.
RA Ohtoko K., Ohkuma M., Moriya S., Kudo T.;
RT "Diverse genes of family 45 cellulase homologues of the symbiotic
RL protists in the hindgut of termite Reticulitermes speratus.";
EMBL: AB045174; BAA98044.1; -
DR HSSP: P43316; 2ENG.
DR InterPro: IPR000334; GH_45.
DR Pfam: PF02015; GLYCOHYDROL_F45; UNKNOWN_1.
DR PROSITE: PS01140; GLYCOSYL_HYDROL_45; UNKNOWN_1.
SQ SEQUENCE 219 AA; 23033 MW; CA295CAD8F93199 CRC64;

Query Match
Best Local Similarity 42.0%; Score 697; DB 14; Length 219;
Matches 123; Conservative 34; Mismatches 50; Indels 8; Gaps 4;

OY 15 LPLVASASGSGSTRYWDCKPSCAMPKAAVSQPVYAC--DANFORLSDFNVSQSGC 72
DB 6 LTFGLSLADSGKTRTYWDCKSGCWBAKADVSKPIDTCAKDDGTTTVAASNDTVKSGC 65

```

```

OY 73 GSAYSCADQTPMAVNDLAVGPAATSIAGSESSWCACALFTTSGPVAGKTVVOSTS 132
DB 66 DCGDGFMCYDQTPMAVNDLSIGFAAAVSGG-ESGACCCCELTFTSGPVNGKKTIVY 124
OY 133 TGGDLGSGNPDIAIPGGGVGIFNGCSSQFGGLP---GAOYGISRRDQCSFPAPLK 188
DB 125 TGGDLGSGNPDIAIPGGGVGIVNGCTAQ--SGAPADGMSRGVGSRSRSCQLPSGLQ 183
OY 189 CGWRFDFWONADNPTFTFOQVOCPAETVARSCKRN 223
DB 184 CGWRFDFWONADNPTFTFNVSQPSLISKTNCRN 218

RESULT 13
OY 09JH88 PRELIMINARY; PRT; 219 AA.
AC 09JH88;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DE 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Family 45 cellulase homologue.
OS unclassified eukaryotes.
OC Eukaryota.
OX NCBI_TaxID=42452;
RN [1]
RP SEQUENCE FROM N.A.
RA Ohtoko K., Ohkuma M., Moriya S., Kudo T.;
RT "Diverse genes of family 45 cellulase homologues of the symbiotic
RL protists in the hindgut of termite Reticulitermes speratus.";
EMBL: AB045175; BAA98045.1; -
DR HSSP: P43316; 2ENG.
DR InterPro: IPR000334; GH_45.
DR Pfam: PF02015; GLYCOHYDROL_F45; UNKNOWN_1.
DR PROSITE: PS01140; GLYCOSYL_HYDROL_45; UNKNOWN_1.
SQ SEQUENCE 219 AA; 23030 MW; 17BFF1344C6D7024 CRC64;

Query Match
Best Local Similarity 42.0%; Score 697; DB 14; Length 219;
Matches 124; Conservative 33; Mismatches 54; Indels 8; Gaps 4;

OY 11 LAALPLVASASGSGSTRYWDCKPSCAMPKAAVSQPVYAC--DANFORLSDFNVSQ 68
DB 2 LLQILFTGLSLAESGKTRTYWDCKSGCWBAKADVSKPIDTCAKDDGTTTVAASNDTVK 61
OY 69 GCGSASVASCADQTPMAVNDLAVGPAATSIAGSESSWCACALFTTSGPVAGKTVV 128
DB 62 GCDGDTGYMCYDQTPMAVNDLSIGFAAAVSGG-ESGACCCCELTFTSGPVNGKKTIV 120
OY 129 QSTTSGDLGSGNPDIAIPGGGVGIFNGCSSQFGGLP---GAOYGISRRDQCSFPAP 184
DB 121 QITNTGDLGSGNPDIAIPGGGVGIVNGCTAQ--SGAPADGMSRGVGSRSRSCQLPSG 179
OY 185 LKPGCGWRFDFWONADNPTFTFOQVOCPAETVARSCKRN 223
DB 180 LKPGCGWRFDFWONADNPTFTFNVSQPSLISKTNCRN 218

RESULT 14
OY 09JH96 PRELIMINARY; PRT; 220 AA.
AC 09JH96;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DE 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Family 45 cellulase homologue.
OS unclassified eukaryotes.
OC Eukaryota.
OX NCBI_TaxID=42452;
RN [1]
RP SEQUENCE FROM N.A.
RA Ohtoko K., Ohkuma M., Moriya S., Kudo T.;

```

